https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

A benchmark dataset for half-hourly evapotranspiration estimation in China from 2000 to 2024

Long Qian[1,2], Xingjiao Yu[1,2], Lifeng Wu[1], Yaokui Cui[3], Zhitao Zhang[2], Junying Chen[2], Sumeng Ye[2], Xuqian Bai[2], Xiaogang Liu[4], Sien Li[5], Rangjian Qiu[6]

[1]School of Soil and Water Conversation, Jiangxi University of Water Resources and Electric Power, Nanchang, 330099, China
[2]College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, 712100, China
[3]Institute of RS and GIS, School of Earth and Space Sciences, Peking University, Beijing 100871, China
[4]Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, China
[5]Center for Agricultural Water Research in China, China Agricultural University, 100083, Beijing, China
[6]State Key Laboratory of Water Resources Engineering and Management, Wuhan University, 430072, Wuhan, China

Correspondence to: Lifeng Wu (china.sw@163.com)

Abstract. Latent heat flux (LE) provides a direct representation of terrestrial evapotranspiration (ET) and plays a critical role in hydrological cycle studies, land surface model development, and the evaluation of remotely sensed evapotranspiration products. Although flux observations based on the eddy covariance technique are widely regarded as essential benchmark data for evapotranspiration estimation, existing ChinaFlux observations are generally limited by short observation periods and extensive data gaps, which substantially constrain their applicability in long-term change analyses and multi-scale studies. To address these limitations, we developed a gap-filling and temporal prolongation framework specifically designed for half-hourly LE and established a continuous ground-based benchmark dataset covering China for the period 2000–2024 based on observations from 50 ChinaFlux sites. The framework is built upon an automated machine learning approach (AutoML-H2O) and integrates ERA5-Land reanalysis data with MODIS vegetation indices, enabling accurate gap-filling within observation periods and reliable prolongation beyond observation intervals. Comprehensive evaluations demonstrate that the AutoML framework achieves high accuracy at the half-hourly scale across different gap-length scenarios, with an overall correlation coefficient (CC) of 0.862 and a root mean square error (RMSE) of 33.75 W m^{-2}, and it substantially outperforms conventional methods under long-gap conditions of 7 d and 30 d. The forward and backward prolongation results show high consistency (CC values of 0.902 and 0.896, respectively) and exhibit robust temporal stability under varying training data lengths. Multi-timescale validations further indicate that the prolonged LE data reasonably reproduce diurnal variations, seasonal cycles, and interannual variability from half-hourly to daily and monthly scales. Comparisons with ChinaFlux observations under strict quality control reveal good consistency across different temporal scales, underlying surface types, and climate zones. SHAP-based interpretability analysis indicates that energy supply consistently dominates LE variability, while vegetation state and water availability modulate their relative importance under different environmental conditions. Overall, we present the first continuous half-hourly ground-based LE benchmark dataset covering China for the

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

period 2000–2024. This dataset provides essential data support for the evaluation of remotely sensed ET products, land surface model validation, and studies of regional water–energy cycles and climate change, and it is freely available via the following repository: https://doi.org/10.5281/zenodo.18194590 (Qian et al., 2026).

1 Introduction

Terrestrial evapotranspiration (ET) is a fundamental component of the land surface hydrological cycle, serving as the primary pathway by which water is transferred from the terrestrial surface to the atmosphere (Tang et al., 2024; Zhang et al., 2025). As a core process linking the atmosphere, hydrosphere, and biosphere, ET plays a pivotal role in regulating exchanges of water, carbon, and energy within the Earth system (Allen et al., 1998; Fisher et al., 2017). From an energy perspective, ET consumes approximately 48–62 % of the net solar radiation received at the land surface, thereby exerting a direct influence on surface cooling effects and cloud–precipitation feedbacks (Trenberth et al., 2009). From a hydrological perspective, more than 60 % of global terrestrial precipitation is returned to the atmosphere through ET, effectively constraining the natural upper limit of renewable freshwater resources (Oki and Kanae, 2006). Consequently, the availability of temporally continuous and quality-controlled ET benchmark data is a prerequisite for advancing our understanding of land–atmosphere interactions, evaluating evapotranspiration products derived from remote sensing and land surface models, and supporting water balance assessments and ecohydrological studies at basin and regional scales. In particular, at the site scale, flux observations based on the eddy covariance technique provide direct measurements of latent heat flux and its variability, and their temporal continuity and internal consistency are critical for model calibration, product validation, and cross-regional comparative analyses (Rodell et al., 2015; Cheng et al., 2021).

However, despite the central role of evapotranspiration in hydrological and ecosystem studies, existing observational approaches remain insufficient to fully meet the demands of regional- and long-term analyses. With the expansion of spatial and temporal scales, land surface models, remote sensing retrievals, and data assimilation systems have become the primary means for generating spatially and temporally continuous ET estimates (Zhang et al., 2022; Yu et al., 2022; Zou et al., 2017). Nevertheless, the accuracy assessment and uncertainty characterization of these approaches rely heavily on calibration and validation against high-quality ground-based observations. Traditional ground-based measurement methods, such as evaporation pans, lysimeters, and the Bowen ratio technique, can provide relatively accurate estimates of ET or surface energy fluxes at the point scale (Allen et al., 2011). However, due to limitations in instrument deployment, maintenance requirements, and long-term operational stability, these methods are unable to form observation networks with broad spatial coverage, long-term continuity, and internal consistency (Xu and Singh, 2005). As a result, ground-based flux observations with high temporal resolution and strong physical consistency are particularly critical for regional and global-scale studies.

Within the existing ground observation systems, the eddy covariance (EC) technique has been widely regarded as the most important data source for constraining and evaluating ET models, as it directly measures turbulent exchanges between the land surface and the atmosphere. EC observations provide half-hourly measurements of latent heat flux (LE) and its driving variables, thereby serving as an essential ground-based benchmark for evapotranspiration research. However, EC flux time series are commonly affected by substantial data gaps and long periods of missing observations due to instrument maintenance, energy balance non-closure, complex meteorological conditions, and data transmission interruptions, with large variations in effective observation lengths among sites. Taking FLUXNET2015 as an example, the average missing rate of hourly LE data is approximately 40 %, exceeding 70 % at some sites, and long gaps lasting more than 30 d account for a considerable fraction of the missing records (Foltýnová et al., 2020; Zhu et al., 2022). Although the marginal distribution sampling (MDS) approach has been widely adopted as the official gap-filling algorithm and performs reasonably well for short gaps, it often fails to accurately reconstruct the magnitude and extremes of LE under long-gap conditions or during periods of extreme environmental variability. Moreover, the overall observation durations of existing flux sites are relatively short, and conventional gap-filling methods do not support temporal prolongation, which further limits the applicability of flux observations in long-term studies.

In China, the aforementioned limitations are also pervasive in ChinaFlux flux observations and are further exacerbated by insufficient site density, limited temporal continuity, and weak long-term consistency. Complex terrain, diverse climate regimes, and strong anthropogenic disturbances make flux tower measurements more susceptible to non-ideal conditions, resulting in frequent data gaps in latent heat flux (LE) time series and pronounced variability in data quality among sites. Although regional flux networks such as ChinaFlux have been established, the number of available flux tower sites in China has long been inadequate for regional- and global-scale ET product development and evaluation. Previous studies have shown that ET product validation over China often relies on fewer than 10 flux tower sites (Guo et al., 2022; Shi et al., 2024; Zuo et al., 2025), which is insufficient to represent the country's highly heterogeneous climate conditions and land surface characteristics. Even in comprehensive assessments based on global flux networks, Chinese sites remain underrepresented; for example, validation frameworks including nearly 200 global sites typically contain only 8–11 sites within China (Liu et al., 2023; Qian et al., 2023). In some data-driven modeling efforts and global ET product training datasets, the proportion of Chinese flux tower samples accounts for less than 2 % of the global total (Elnashar et al., 2021; Lu et al., 2021), substantially limiting model applicability and generalization capability over China.

Although FLUXNET2015 has provided critical support for global evapotranspiration research (Jung et al., 2019; Qian et al., 2024a), its spatial coverage and temporal continuity within China remain limited. Recent studies indicate that global ET benchmark datasets derived from FLUXNET2015 include only a small number of Chinese sites and lack long-term continuous half-hourly flux time series for this region (Li et al., 2025). Meanwhile, several systematic evaluation studies focusing on China have primarily been conducted at the monthly scale (Liu et al., 2025), which hampers the characterization of ET dynamics at diurnal and short temporal scales. Under conditions of limited site availability and discontinuous observation records, discrepancies among different ET products over China are further amplified, with annual-scale estimates differing by up to nearly 50 % and exhibiting more pronounced regional contrasts in temporal trends (Cai et al., 2024). Against this background, there is an urgent need to systematically perform data gap-filling and temporal prolongation based on existing ChinaFlux observations and to establish a Chinese regional ground-based evapotranspiration benchmark

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

dataset with extended temporal coverage and unified quality control standards. Such a dataset is essential for supporting the evaluation of remotely sensed ET products, the development of land surface and data-driven models, and studies of regional hydrological and ecohydrological processes.

Although model-based and remote sensing approaches are widely used for regional-scale evapotranspiration estimation, gap-filling and time-series reconstruction of flux tower latent heat flux (LE) observations still primarily rely on statistical and empirical methods. Among these, marginal distribution sampling (MDS) has been widely adopted as a standard approach due to its simplicity and reasonable performance for short gaps (Reichstein et al., 2005; Pastorello et al., 2020). However, because evapotranspiration is jointly controlled by energy availability, water supply, and vegetation dynamics, its responses are highly nonlinear and context dependent, which limits the ability of MDS—based on meteorological similarity assumptions—to represent complex causal relationships (Chen et al., 2012).Previous studies have shown that even with the inclusion of additional variables such as soil moisture or vegetation indices, MDS tends to excessively smooth LE variability under extreme conditions or pronounced interannual variability, leading to substantial errors when continuous gaps exceed approximately 15 d (Moffat et al., 2007; Kang et al., 2019; Zhu et al., 2022). Other conventional approaches, including mean diurnal variation, lookup table, and regression-based methods, exhibit similar limitations and are prone to systematic biases when data gaps are extensive or environmental conditions change markedly (Qian et al., 2024b). Moreover, these methods are restricted to gap-filling within observation periods and do not support temporal extension of flux time series, thereby constraining their application in long-term studies.Against this background, data-driven approaches have emerged as a promising alternative for LE gap-filling and temporal prolongation at flux tower sites. By learning nonlinear relationships between LE and its meteorological and surface drivers directly from observations, these methods offer greater flexibility in handling long data gaps and complex environmental conditions when reference variables are continuously available (Kim et al., 2020; Zhu et al., 2022; Guo et al., 2022).Nevertheless, conventional machine learning models typically rely on manually specified model structures and hyperparameters, while deep learning approaches are often sensitive to sample size and data distribution, which can introduce instability when the number of available flux sites is limited (Khan et al., 2021). Automated machine learning (AutoML), by systematically exploring model architectures and optimizing hyperparameters, reduces reliance on subjective experience while enhancing model robustness. When combined with reanalysis data featuring high temporal and spatial continuity, AutoML not only enables effective flux gap-filling but also provides a feasible approach for extending observation-based time series beyond measurement periods, thereby laying a methodological foundation for constructing long-term, continuous, site-level evapotranspiration benchmark datasets.

To address the widespread issues of extensive data gaps, inconsistent observation periods, and long-term sample scarcity in flux tower latent heat flux (LE) observations over China, we developed a site-level gap-filling and temporal prolongation framework for LE and, based on this framework, established a continuous evapotranspiration benchmark dataset for Chinese flux tower sites covering the period 2000–2024. Following rigorous quality control and site selection criteria, this study integrates observations from 50 ChinaFlux flux tower sites whose spatial distribution spans the major climate zones and representative underlying surface types across China, providing substantially improved site density and regional

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

representativeness compared with previous studies. Methodologically, the framework is built upon an automated machine learning approach (AutoML-H2O) and combines reanalysis meteorological data with remotely sensed vegetation information to model the nonlinear relationships between LE and its driving factors at the site scale. This approach enables both high-accuracy gap-filling within observation periods and reasonable temporal prolongation beyond measurement intervals, resulting in continuous half-hourly LE time series. Through a unified data processing workflow and consistent modeling strategy, LE data from all sites are harmonized in terms of temporal resolution and quality control standards, effectively alleviating the constraints imposed by long continuous data gaps and heterogeneous observation periods in the original measurements. To systematically assess the reliability and applicability of the constructed dataset, we designed a multi-level evaluation framework that includes gap-filling performance assessments under different gap-length scenarios, consistency tests between forward and backward temporal prolongation, and analyses of long-term temporal stability. These evaluations were further conducted across different underlying surface types and climate zones. The resulting dataset provides a temporally continuous, quality-controlled, and physically consistent ground-based evapotranspiration benchmark for China, offering essential data support for the evaluation of remotely sensed ET products, validation of reanalysis datasets and land surface process models, and studies of regional hydrological and ecosystem processes, while also establishing a robust data foundation for analyzing water–energy–carbon cycle dynamics over China under a changing climate.

2 Data and methodology

2.1 ChinaFlux

This study is based on eddy covariance (EC) observations from the ChinaFlux network, the primary long-term observation system for land–atmosphere interactions in China. Since 2002, ChinaFlux has been progressively established and operated across the country using standardized instrument configurations, observation protocols, and data processing procedures. The network systematically measures carbon, water, and energy fluxes over terrestrial ecosystems, together with accompanying meteorological variables, and provides one of the most important ground-based observational foundations for regional- and national-scale evapotranspiration (ET) studies in China. Within the ChinaFlux network, all flux towers employ the standard EC technique to continuously measure latent heat flux (LE). Raw high-frequency measurements are processed through a unified workflow to generate half-hourly flux products. This processing typically includes three-dimensional coordinate rotation, frequency response correction, Webb–Pearman–Leuning (WPL) correction, outlier removal, friction velocity (u*) filtering, and evaluation of energy balance closure, following quality control (QC) standards compatible with both ChinaFlux and FLUXNET. Previous studies have demonstrated that long-term observations from ChinaFlux sites are generally of reliable quality, with energy balance closure levels falling within internationally accepted ranges, thereby meeting the basic requirements for use as regional ET benchmark data.

In total, this study integrates and selects 50 ChinaFlux flux tower sites (Fig. 1a), covering six major underlying surface types in China, including grassland (16 sites), shrubland (2 sites), cropland (9 sites), forest (13 sites), desert (5 sites), and wetland

(5 sites). These sites also span seven major climate zones: the Northeast Temperate Subhumid Zone (NETSZ, 4 sites), Inner Mongolia Temperate Semiarid Zone (IMSZ, 5 sites), Northern Temperate Subhumid Warm Zone (NTSWZ, 9 sites), Southeast Subtropical Humid Zone (SESHZ, 7 sites), Southern Tropical Humid Zone (STHZ, 3 sites), Northwest Desert Arid Zone (NWDAZ, 10 sites), and Qinghai–Tibet Plateau Semiarid Zone (QTPSZ, 12 sites). The selected sites exhibit substantial heterogeneity in climatic conditions, vegetation types, and hydrothermal regimes, enabling a robust representation of the spatial variability of evapotranspiration processes across China. Detailed information on site locations, underlying surface types, climate zone classifications, and observation periods is provided in Table A1.

Given the limited number of flux tower sites and the scarcity of long-term continuous records in China, a relatively inclusive site selection strategy was adopted to maximize the use of available information while ensuring basic data reliability. Specifically, selected sites were required to meet the following criteria: (1) an effective observation period of at least 2 years, and (2) no fewer than 10 000 half-hourly LE records available for model training, ensuring a stable basis for gap-filling and temporal prolongation. Among the final set of sites, 40 provide quality-controlled original LE observations, while the remaining 10 sites offer continuous LE time series that have already been gap-filled within their observation periods. Given the overall scarcity of ChinaFlux data, these sites were retained to allow users flexibility in data selection according to specific research objectives. The temporal coverage and data gap characteristics of the selected sites are summarized in Fig. 1b–d, which illustrate the proportion of missing LE data, interannual variations in the number of available sites, and the distribution of observation period lengths, respectively. All selected sites provide half-hourly LE data, forming the foundation for subsequent gap-filling, temporal prolongation, and the construction of a continuous dataset spanning 2000–2024. By integrating these sites, this study assembles one of the most comprehensive flux tower–based evapotranspiration observation datasets for China to date in terms of site number and coverage of climate and underlying surface types, providing essential support for methodological evaluation, long-term consistency analyses, and the development of a regional ET benchmark dataset for China.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

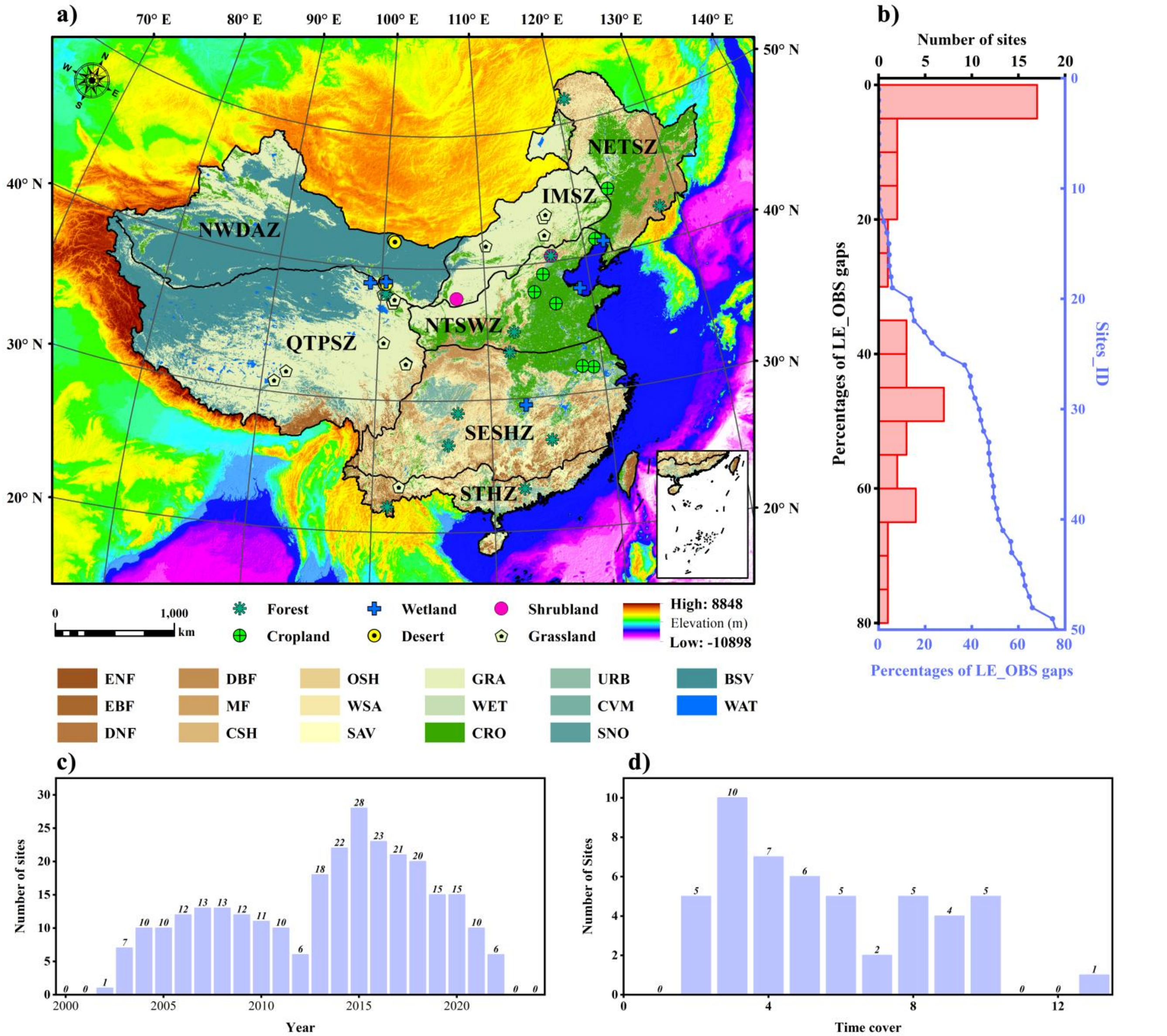


Figure 1: Spatial distribution of the selected ChinaFlux sites (a) and data coverage characteristics of half-hourly LE observations (b–d). Panel (b) shows the distribution of LE data gap percentages across sites, panel (c) presents the number of available sites by year, and panel (d) summarizes the length of observation periods for all sites.

2.2 ERA5-Land data

To support gap-filling and temporal prolongation of flux tower latent heat flux (LE) observations, ERA5-Land reanalysis data were selected as site-scale meteorological and hydrological driving variables. ERA5-Land is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and provides globally continuous coverage with high spatiotemporal consistency, offering a uniform background for flux reconstruction across different climate zones and land surface conditions. Hourly ERA5-Land data were extracted at each site location from the ECMWF/ERA5_LAND/HOURLY dataset using Google Earth Engine (GEE; https://code.earthengine.google.com/, last access: 11 November 2025). Bilinear interpolation was applied to derive representative values at the site scale, and all timestamps were converted from Coordinated Universal Time (UTC) to local time to ensure consistency with flux tower observations. To match the half-hourly temporal resolution of the EC measurements, hourly ERA5-Land variables were resampled to a half-hourly scale using linear interpolation. The ERA5-Land variables used in this study include latent heat flux (LE), surface air pressure (PA), precipitation (Pre), relative humidity (RH), net solar radiation (Rn), downwards solar radiation (Rs), surface runoff (Runoff), volumetric soil water in the 0–7 cm layer (SMC_1) and volumetric soil water in the 7–28 cm layer (SMC_2), air temperature (Temp), vapor pressure deficit (VPD), and 10 m wind speed (WS). These variables collectively characterize key controls on evapotranspiration, including surface energy balance, atmospheric moisture and aerodynamic conditions, and soil moisture constraints.

2.3 MODIS data

To complement the meteorological drivers and better represent vegetation conditions, remotely sensed products from the Moderate Resolution Imaging Spectroradiometer (MODIS) were used to derive the normalized difference vegetation index (NDVI) and leaf area index (LAI). NDVI data were obtained from the MODIS MOD13Q1 (Terra) and MYD13Q1 (Aqua) products, while LAI data were derived from the MODIS MOD15A2H and MYD15A2H products. All products were converted to physical values using the corresponding scale factors provided in the official documentation. NDVI and LAI data were extracted at the site scale using GEE (https://code.earthengine.google.com/, last access: 11 November 2025), with bilinear interpolation applied for spatial sampling. The spatial resolutions of the NDVI and LAI products are 250 m and 500 m, respectively. Given the relatively low temporal resolution of MODIS products, NDVI and LAI were assumed to remain constant within each compositing period, and the corresponding values were assigned uniformly to all half-hourly (and daily) time steps within that period to ensure temporal alignment with the half-hourly LE time series.

It should be noted that MODIS NDVI and LAI products are available from 18 February 2000 at 08:00 (local time). For periods prior to this date, NDVI and LAI data are unavailable. In these cases, vegetation indices were treated as missing values and were not included as predictors in model training and prediction. Given that the AutoML framework is primarily based on tree-based algorithms capable of handling missing values, LE estimation during this early period relied on meteorological and energy-related drivers.

2.4 Generation of gap-filling

To construct temporally continuous latent heat flux (LE) time series suitable for long-term consistency analyses, a data-driven gap-filling framework was implemented to reconstruct missing LE observations at the flux tower scale. Given the substantial heterogeneity among flux sites in terms of climate background, underlying surface type, and observation length, a site-specific modeling strategy was adopted. Each flux tower was treated as an independent modeling unit, for which models

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

were separately trained, evaluated, and applied to generate seamless half-hourly LE time series over the study period. During model development and evaluation, the original LE observations were first subjected to quality screening, and only reliable records were retained for model training. To systematically assess model performance under different gap conditions, multiple artificial gap scenarios were constructed along the time dimension by deliberately removing data segments of varying lengths from the complete observation records. This design enables a robust evaluation of gap-filling performance across different temporal scales. In addition, conventional approaches widely used in flux studies were included as benchmark methods, allowing quantitative comparisons with data-driven approaches in terms of accuracy and stability. The overall technical workflow and evaluation framework are illustrated in Fig. 2.

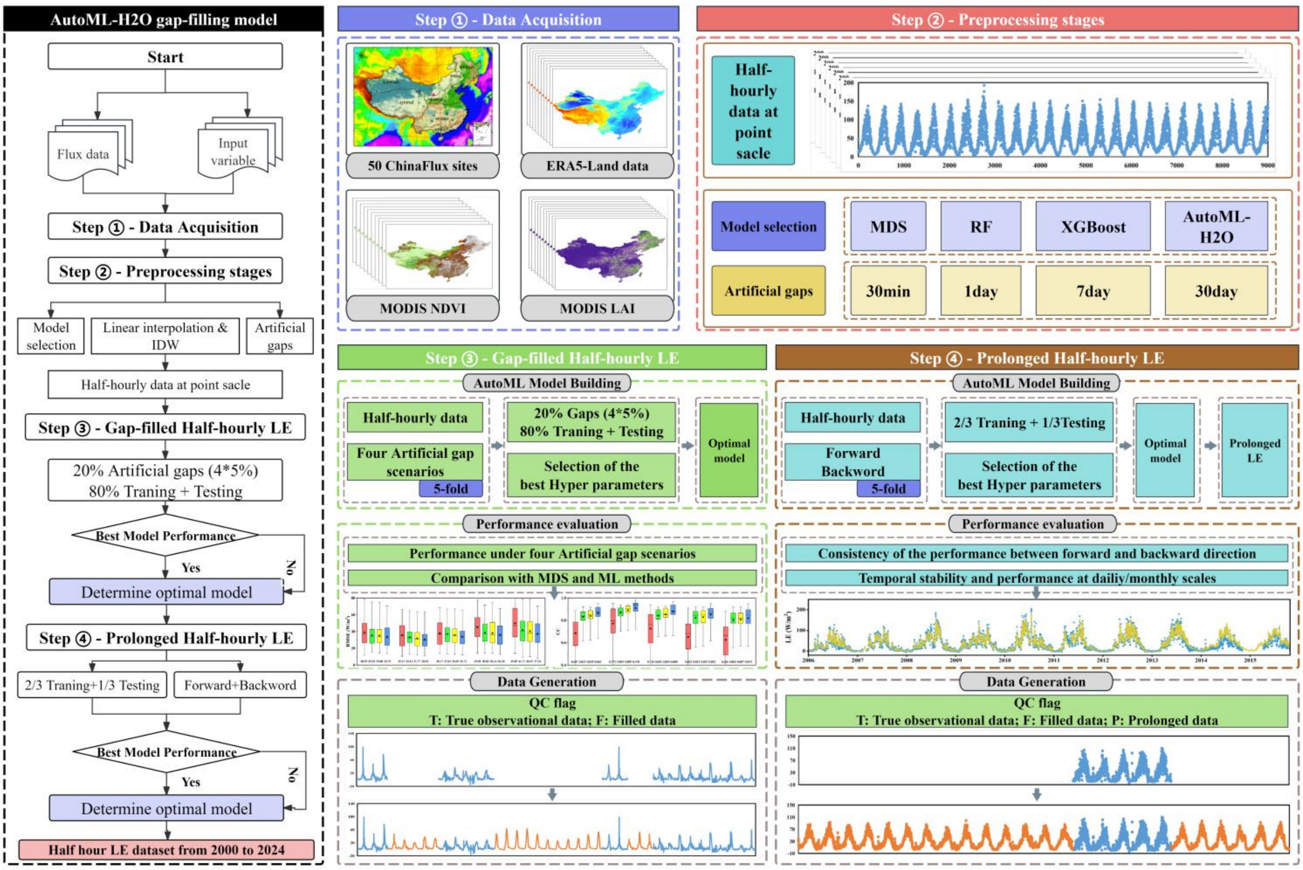


Figure 2: Flowchart and schematic of the gap-filling and prolongation framework for LE_OBS data.

2.4.1 AutoML-H2O

H2O AutoML was adopted as the core modeling tool for LE gap-filling in this study. Unlike conventional machine learning approaches that require manual specification of model types and hyperparameters—often leading to high tuning costs and reduced transferability across sites—AutoML automatically performs model selection, hyperparameter optimization, and model ranking within a predefined search space. This strategy improves modeling efficiency and robustness across multiple sites while maintaining reproducibility. For each flux tower site, an independent AutoML regression model was constructed, with site-level LE observations as the target variable and a set of multi-source environmental drivers derived from ERA5-Land reanalysis data and MODIS vegetation indices as input features. These predictors were used to capture the nonlinear relationships between LE and its meteorological, hydrological, and vegetation controls. During the AutoML search process, multiple commonly used regression algorithms, including tree-based models and their ensemble variants, were evaluated, and the optimal model was automatically selected based on validation performance.

In the training and validation stage, the valid observations at each site were randomly divided into a training set (approximately 80 %) and a testing set (approximately 20 %). Model performance was evaluated through repeated random splits combined with five-fold cross-validation to reduce uncertainties associated with data partitioning and to mitigate overfitting. After model selection and hyperparameter optimization, the optimal model for each site was retrained using all available high-quality LE observations and subsequently applied to fill missing LE values. As a result, each site was assigned a site-specific optimal gap-filling model consistent with its local observation conditions, providing a reliable foundation for subsequent temporal prolongation of LE time series.

2.4.2 Artificial gap scenarios

In flux tower latent heat flux (LE) observations, data gaps vary substantially in duration, ranging from single missing time steps to continuous gaps lasting several weeks or longer. To systematically evaluate model performance across different missing-data scales, four artificial gap scenarios were constructed within the observation period of each site, corresponding to continuous gap lengths of 30 min, 1 d, 7 d, and 30 d. For each scenario, the artificially removed data accounted for approximately 5 % of the total valid observations at the site, ensuring comparability of evaluation results across gap-length conditions. Artificial gaps were generated using a sliding-window approach to create continuous missing segments. Only time windows in which the proportion of valid original observations exceeded 50 % were considered eligible, thereby avoiding the introduction of artificial gaps within periods of inherently poor data quality. Artificial gaps generated under different scenarios did not overlap in time. For the 30 min scenario, single half-hourly observations were randomly removed, with the additional constraint that valid observations were present immediately before and after the removed record to preserve the basic continuity of the time series. During model training and validation, the remaining data after artificial gap removal were randomly divided into a training set (80 %) and a validation set (20 %). Five-fold cross-validation was applied during training to evaluate model performance. For each data split, the model parameters yielding the best validation performance were retained and subsequently applied to fill the corresponding artificial gaps, allowing assessment of gap-filling accuracy and stability under different missing-data scales. This procedure was conducted independently for each site and repeated 20 times using different random data splits and gap realizations to reduce the influence of randomness. After completing the performance evaluation across all gap scenarios, the optimal model for each site was retrained using all available high-quality LE observations and applied to fill missing data within the observation period. This process resulted in

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

seamless half-hourly LE time series, which serve as the baseline dataset for subsequent temporal prolongation and multi-scale analyses.

To facilitate comparative evaluation of different gap-filling approaches, the traditional marginal distribution sampling (MDS) method and two commonly used machine learning algorithms, random forest (RF) and extreme gradient boosting (XGBoost), were implemented as benchmark methods. It should be noted that most ChinaFlux sites do not provide complete sets of meteorological driving variables (e.g., shortwave radiation, air temperature, and vapor pressure deficit). Therefore, ERA5-Land variables were consistently used as reference inputs for all benchmark methods to ensure that different algorithms were compared under identical information conditions. This practice has been widely adopted in flux data gap-filling studies and provides a reasonable representation of site-scale meteorological backgrounds when in situ measurements are unavailable. The MDS method performs gap-filling based on meteorological similarity assumptions, whereas RF and XGBoost predict LE by learning statistical relationships between LE and multi-source environmental drivers. All benchmark methods were implemented under the same artificial gap scenarios and data partitioning schemes as AutoML, ensuring fairness and reproducibility in performance comparisons.

In addition, four commonly used performance metrics were employed to quantify gap-filling performance, including the correlation coefficient (CC), root mean square error (RMSE, W m^{-2}), and percentage bias (PBias, %). The definitions of CC and RMSE follow Li et al. (2025), while PBias was calculated according to Qian et al. (2023).

2.4.3 Prolonged hourly LE data

After completing seamless gap-filling within the observation period, the latent heat flux (LE) time series at each flux tower site were further temporally prolonged at the half-hourly scale to construct continuous records covering a unified time span. Temporal prolongation was also implemented within the AutoML-H2O framework, using the same model architecture, input feature set, and training strategy as described in Sect. 2.4.1 to ensure methodological consistency. For each site, a site-specific prolongation model was independently trained using the available LE observations as supervisory information. Five-fold cross-validation was applied to identify the optimal model configuration and hyperparameter combination. Temporal prolongation was performed in two directions at the half-hourly scale. For a site with an observation period spanning a specific interval (e.g., 2006–2015), the period preceding the observations (e.g., 2000–2005) was defined as backward prolongation, whereas the period following the observations (e.g., 2016–2022) was defined as forward prolongation. Under identical data and modeling conditions, the prolongation performance in both temporal directions is expected to be comparable. To evaluate the consistency of prolongation results across temporal directions, a symmetric data partitioning strategy was adopted. For backward prolongation, the last two-thirds of the observed time series were used for model training, while the first one-third served as the testing set. Conversely, for forward prolongation, the first two-thirds of the data were used for training and the remaining one-third for testing. Model performance on the testing sets under the two scenarios was compared to assess the directional consistency of the prolongation approach.

In addition, to examine the temporal stability of model performance as the length of the prolongation period increases, two representative training-length scenarios were designed. For sites with observation periods of at least 6 years, the first 6 years of data were used for training and the remaining years for testing. For sites with observation periods of at least 3 years, the first 3 years were used for training, with subsequent years reserved for testing. By comparing prolongation performance under different training-length conditions, the long-term stability and reliability of the model during temporal extension were systematically evaluated. All prolongation and validation procedures were conducted independently at the site scale. After completing consistency and stability assessments, the final model for each site was retrained using all available high-quality LE observations and applied to periods outside the observation interval, generating continuous half-hourly LE time series covering 2000–2024. These prolonged datasets provide the basis for subsequent construction and analysis of multi-temporal-scale LE products.

2.5 Generation of LE data

Based on the gap-filling and temporal prolongation procedures described above, seamless site-level half-hourly latent heat flux (LE) time series were finally constructed covering the period from 00:00 on 1 January 2000 to 23:30 on 31 December 2024. The resulting dataset is temporally continuous, processed using a unified workflow across all sites, and explicitly distinguishes data generated through different procedures. At the half-hourly scale, LE records are assigned one of three quality flags according to their data source. The flag T denotes half-hourly LE values directly derived from original flux tower observations. The flag F represents half-hourly LE values obtained by gap-filling missing observations within the measurement period using the AutoML-H2O framework. The flag P indicates half-hourly LE values generated through temporal prolongation beyond the observation period based on site-level models. These quality flags enable users to flexibly select datasets consisting solely of observed data or to include gap-filled and prolonged records depending on specific research objectives.

On this basis, the half-hourly LE data were further aggregated to generate daily, monthly, and annual LE datasets spanning 2000–2024. Daily LE values were obtained by aggregating half-hourly records, while monthly and annual values were subsequently derived from the corresponding daily datasets. LE products at all temporal scales share the same time coverage and quality flagging scheme as the half-hourly data, ensuring consistency and traceability across multi-temporal-scale products.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

3 Result

3.1 Evaluation of half-hourly gap-filled LE data

3.1.1 Gap-filling performance under four artificial gap scenarios

Under the four artificial gap-length scenarios (30 min, 1 day, 7 day, and 30 day), the evaluated methods exhibit pronounced differences in half-hourly LE gap-filling performance (Fig. 3). Overall, AutoML consistently achieves the best and most stable performance across all gap scales, with an average correlation coefficient (CC) of 0.862 and a root mean square error (RMSE) of 33.75 W m^{-2}. This performance substantially exceeds that of the conventional marginal distribution sampling (MDS) method (CC = 0.687) and is accompanied by a more concentrated bias distribution. For the shortest gap scenario (30 min), all methods are generally able to reconstruct short-term LE variability. Nevertheless, AutoML still yields the highest correlation (CC = 0.910) and the lowest error (RMSE = 30.52 W m^{-2}), reducing RMSE by 4.75 W m^{-2} relative to MDS. The performances of RF and XGBoost fall between AutoML and MDS, with overall biases close to zero. As gap length increases, performance differences among methods become progressively more pronounced.

Under the 1 day and 7 day scenarios, MDS performance deteriorates continuously, whereas AutoML maintains relatively high accuracy and stable bias control, indicating strong adaptability to moderate-length continuous gaps. The divergence among methods is most evident under the most extreme scenario of 30 d continuous gaps. In this case, AutoML still achieves a CC of 0.819 and an RMSE of 37.26 W m^{-2}, whereas the RMSE of MDS increases markedly to 49.85 W m^{-2}. Although RF and XGBoost outperform MDS under this scenario, their overall accuracy and stability remain inferior to those of AutoML.Overall, gap-filling performance exhibits a clear dependence on gap length. The MDS approach is only suitable for short gaps and shows rapid performance degradation as gap length increases. In contrast, AutoML maintains higher correlations, lower errors, and more stable bias characteristics across all gap-length scenarios, with particularly pronounced advantages under long continuous gaps. These results indicate that AutoML is better suited for constructing long-term, continuous half-hourly LE time series.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

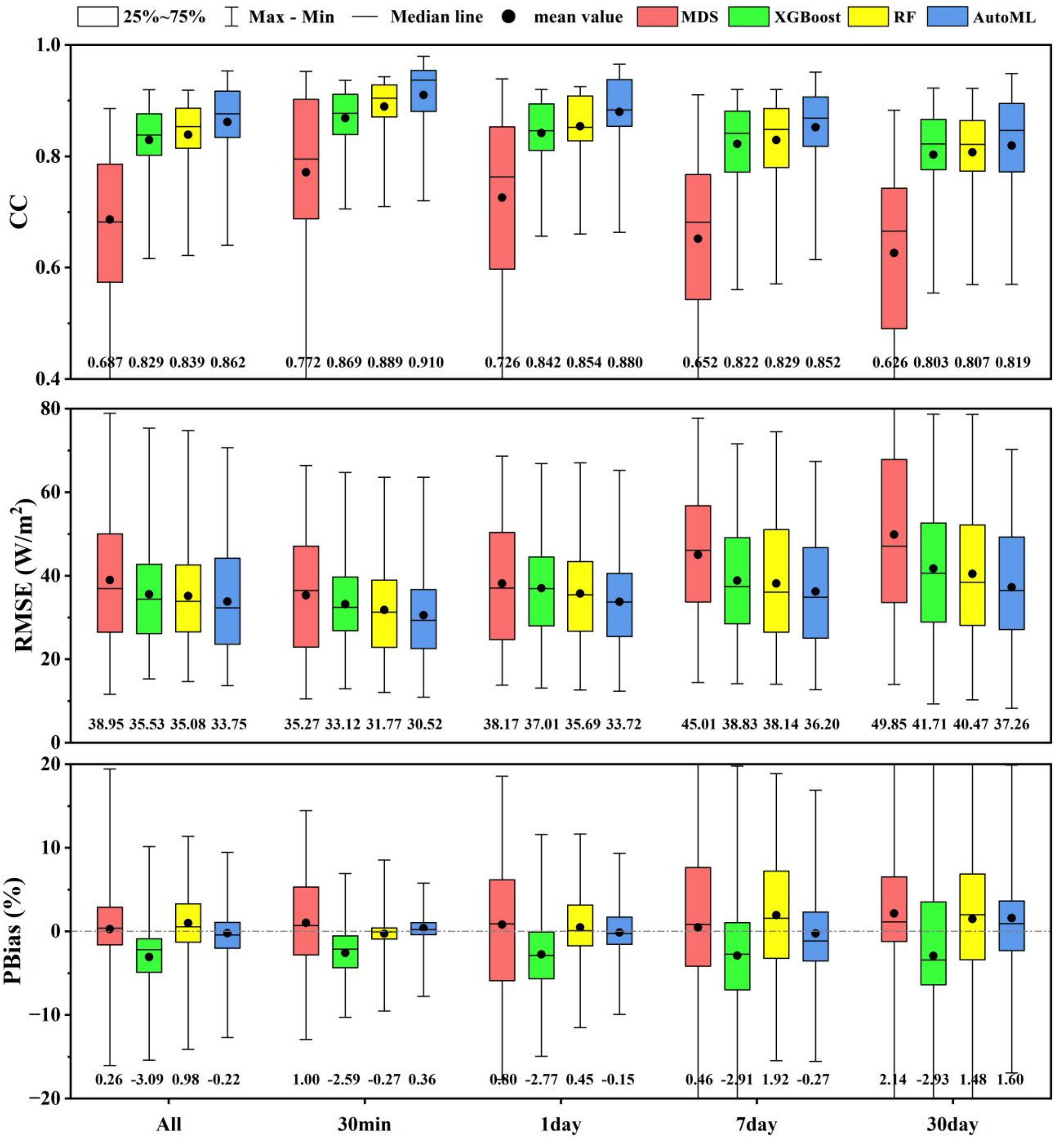


Figure 3: Comparison of half-hourly LE gap-filling performance among different methods under various artificial gap scenarios. Values shown in the figure represent the mean values of the corresponding performance metrics..

Across different underlying surface types (Fig. 4a–c), all methods exhibit a common pattern of decreasing gap-filling performance with increasing gap length, with the most pronounced degradation occurring under the 30 day continuous gap scenario. In terms of overall differences among underlying surface types, cropland and wetland show the best gap-filling performance, maintaining relatively high correlations and low errors across all gap-length scenarios (e.g., AutoML achieves overall CC values of 0.885 and 0.894 for cropland and wetland, respectively). Forest and grassland exhibit intermediate

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

performance, with gap-filling accuracy and stability falling between the high- and low-performing underlying surface types. In contrast, desert and shrubland represent the most challenging underlying surface types for gap-filling. Except for AutoML, the other methods generally show larger errors, pronounced systematic biases, and reduced stability under these conditions. AutoML maintains a relatively concentrated PBias distribution and lower RMSE even under these unfavorable conditions, indicating stronger robustness. Results across different climate zones (Fig. 4d–f) reveal a consistent dependence on gap length, with all methods experiencing the most substantial performance degradation under the 30 day scenario. In terms of regional differences, gap-filling performance is relatively better in the Northwest Desert Arid Zone (NWDAZ) and the Qinghai–Tibet Plateau Semiarid Zone (QTPSZ), characterized by higher correlations and lower errors (e.g., AutoML achieves an overall RMSE of 27.96 W m^{-2} in NWDAZ). In contrast, the Southeast Subtropical Humid Zone (SESHZ), Southern Tropical Humid Zone (STHZ), and Northern Temperate Subhumid Warm Zone (NTSWZ) exhibit lower overall accuracy, with higher error levels and greater bias variability. This pattern reflects the increased uncertainty of evapotranspiration processes under humid and subtropical climate conditions.

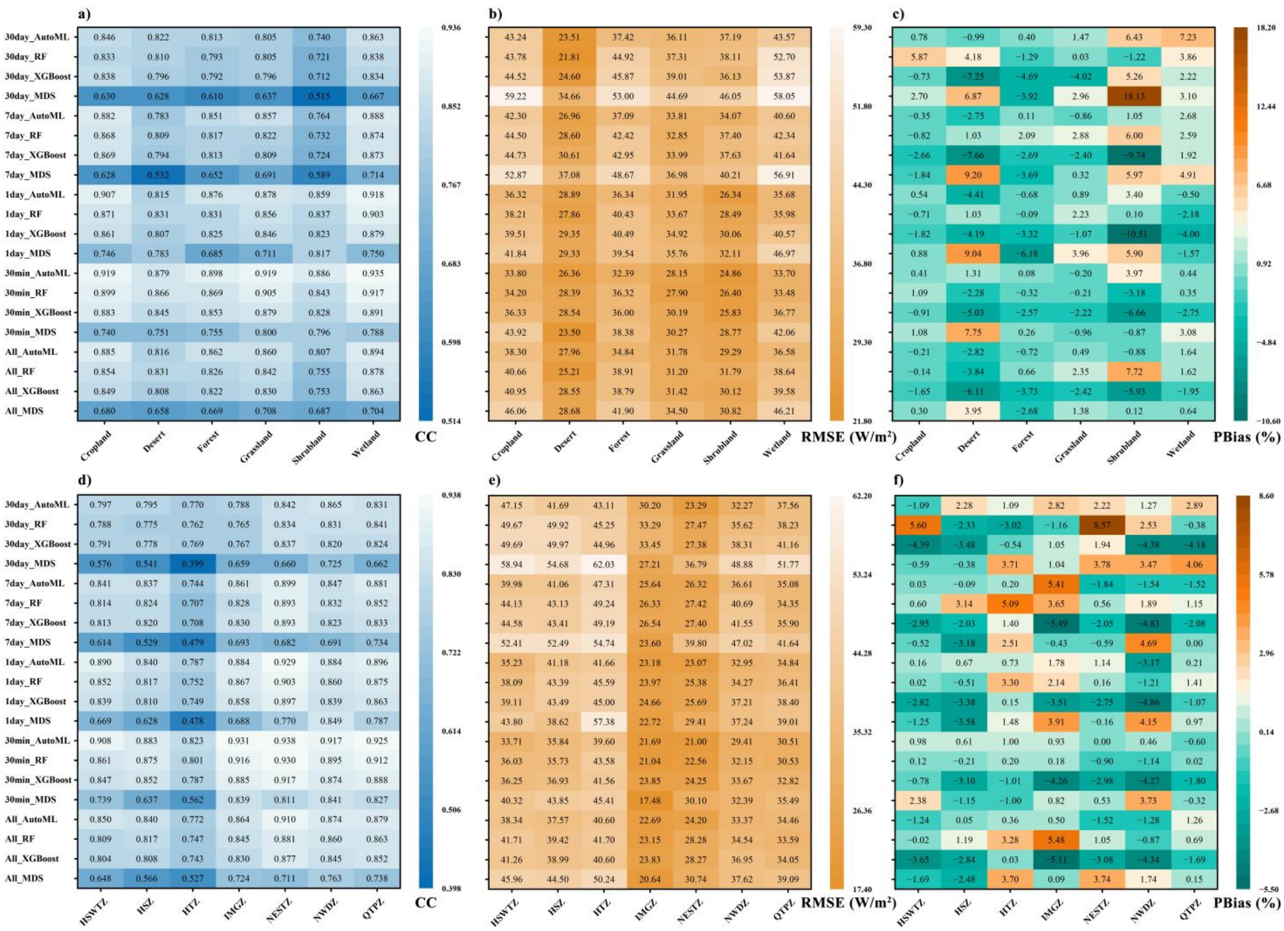


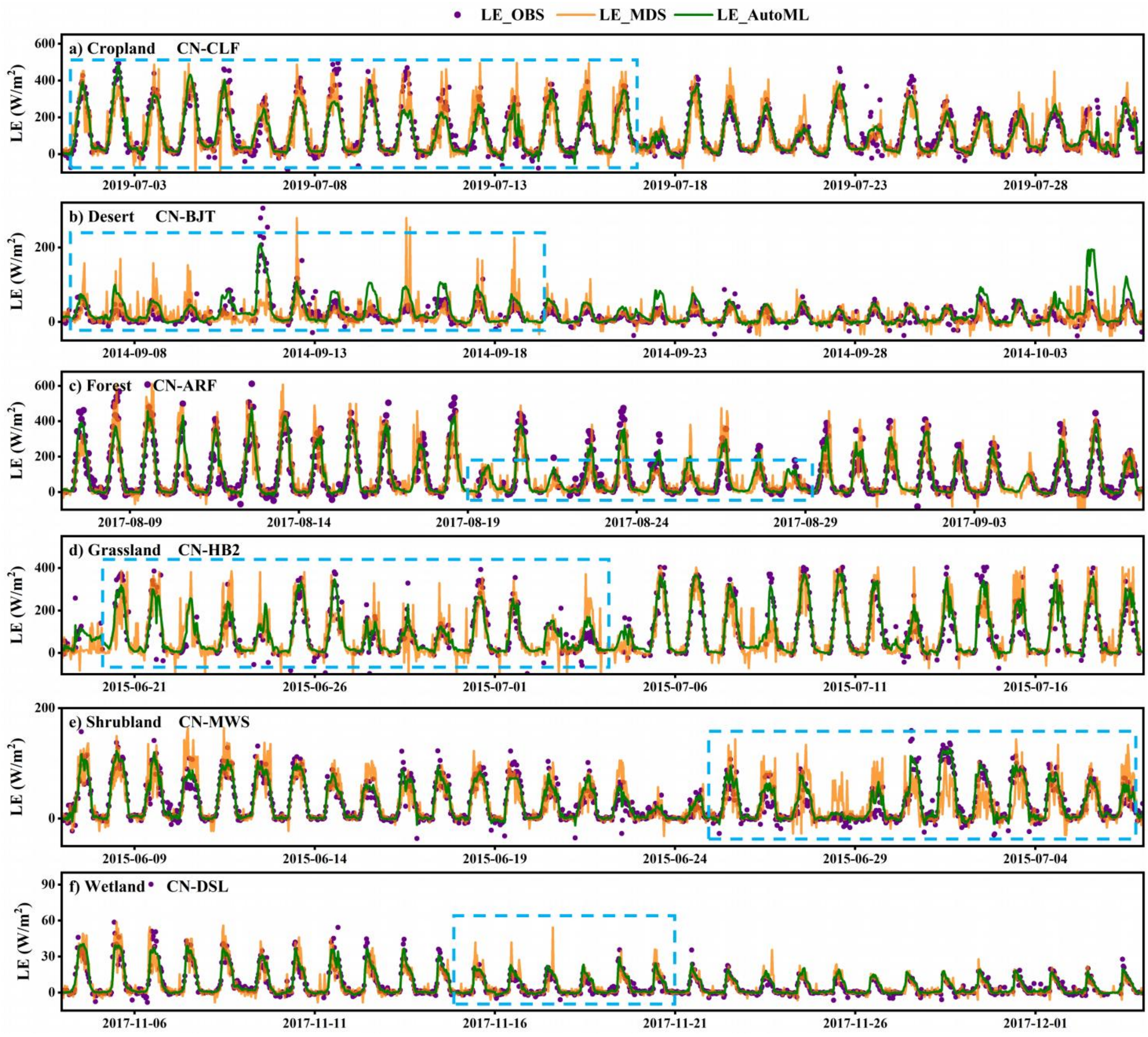
Figure 4: Performance of different methods across underlying surface types and climate zones under varying artificial gap scenarios.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

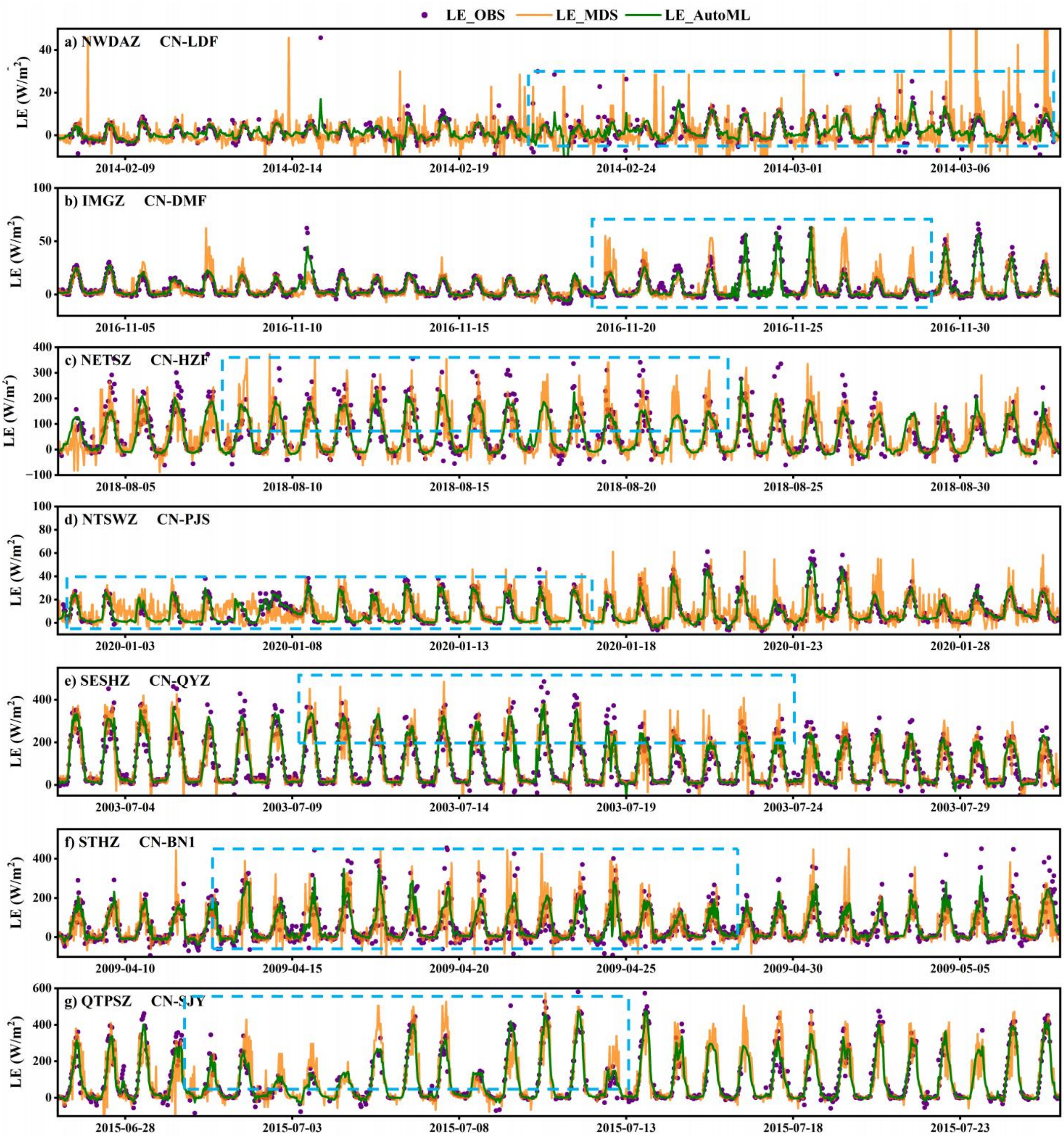
3.1.2 Examples of gap-filled data under artificial 30 day gap-length scenario

Under the artificially imposed 30 day continuous gap scenario, the evaluated gap-filling methods exhibit pronounced differences in their ability to reconstruct half-hourly LE time series (Figs. 5 and 6). Across different underlying surface types, AutoML consistently reproduces the diurnal cycles and amplitude characteristics of LE at cropland, forest, grassland, and wetland sites. Even during extended missing periods, the gap-filled results remain in good agreement with the observed time series. In contrast, MDS frequently exhibits pronounced overestimation or underestimation across multiple underlying surface types, leading to weakened or distorted diurnal structures; this issue is particularly evident at desert and shrubland sites. Results across different climate zones show a similar pattern. In arid and high-altitude regions such as the NWDAZ and QTPSZ, AutoML maintains coherent and continuous temporal patterns during the gap period, whereas MDS displays larger fluctuations and reduced stability. In humid and subtropical climate zones, including NETSZ, NTSWZ, SESHZ, and STHZ, overall uncertainty increases; nevertheless, AutoML is still able to capture the primary temporal structure of LE, while MDS tends to amplify noise and attenuate diurnal signals. Overall, these representative examples demonstrate that under long continuous gap conditions, AutoML more effectively preserves the physical plausibility and temporal consistency of LE time series across different underlying surface types and climate zones. This capability substantially reduces the risk of abnormal fluctuations and structural distortions in the reconstructed LE records.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026



395 **Figure 5: Comparison of gap-filled half-hourly LE time series produced by the AutoML and MDS methods across different underlying surface types under the artificial 30 day gap scenario (the blue dashed boxes indicate scenarios where the MDS gap-filling results are significantly biased).**



400 **Figure 6: Comparison of gap-filled half-hourly LE time series produced by the AutoML and MDS methods across different climate zones under the artificial 30 day gap scenario (the blue dashed boxes indicate scenarios where the MDS gap-filling results are significantly biased).**

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

3.2 Evaluation of half-hourly prolonged LE

3.2.1 Consistency between forward and backward prolongation

Figure 7 compares the overall consistency of half-hourly latent heat flux (LE) under forward and backward temporal prolongation and examines their performance across different underlying surface types and climate zones. Overall, the prolongation results in the two temporal directions are highly consistent in a statistical sense, indicating that the proposed approach is largely insensitive to the direction of temporal extrapolation and exhibits good stability. The overall scatter distributions show that both forward and backward prolongation results agree well with observations. The backward prolongation yields a correlation coefficient (CC) of 0.896 and an RMSE of 35.89 W m^{-2}, while the corresponding values for forward prolongation are 0.902 and 35.33 W m^{-2}, respectively. The differences between the two directions are small, and their overall accuracies are at comparable levels.

Across different underlying surface types, cropland, forest, grassland, and wetland exhibit high correlations under both forward and backward prolongation, with CC values generally close to or exceeding 0.89 and only minor directional differences. Among these, wetland shows high CC values in both directions but relatively larger RMSE, reflecting the larger amplitude of LE variability for this underlying surface type. In contrast, desert and shrubland display slightly lower prolongation accuracy and somewhat larger differences between forward and backward results, suggesting that under land cover conditions characterized by limited observations or higher variability, the influence of extrapolation direction becomes relatively more pronounced. Results across different climate zones further support these findings. NETSZ, IMSZ, QTPSZ, and NWDAZ maintain high consistency between forward and backward prolongation, with most CC values exceeding 0.90. Although SESHZ and STHZ exhibit relatively higher RMSE, the differences between the two temporal directions remain limited, indicating that the associated uncertainty primarily arises from climatic conditions rather than the direction of temporal prolongation. Overall, forward and backward temporal prolongation demonstrate a high degree of consistency in terms of overall accuracy and performance across underlying surface types and climate zones, with only minor differences observed under a limited number of high-uncertainty scenarios.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

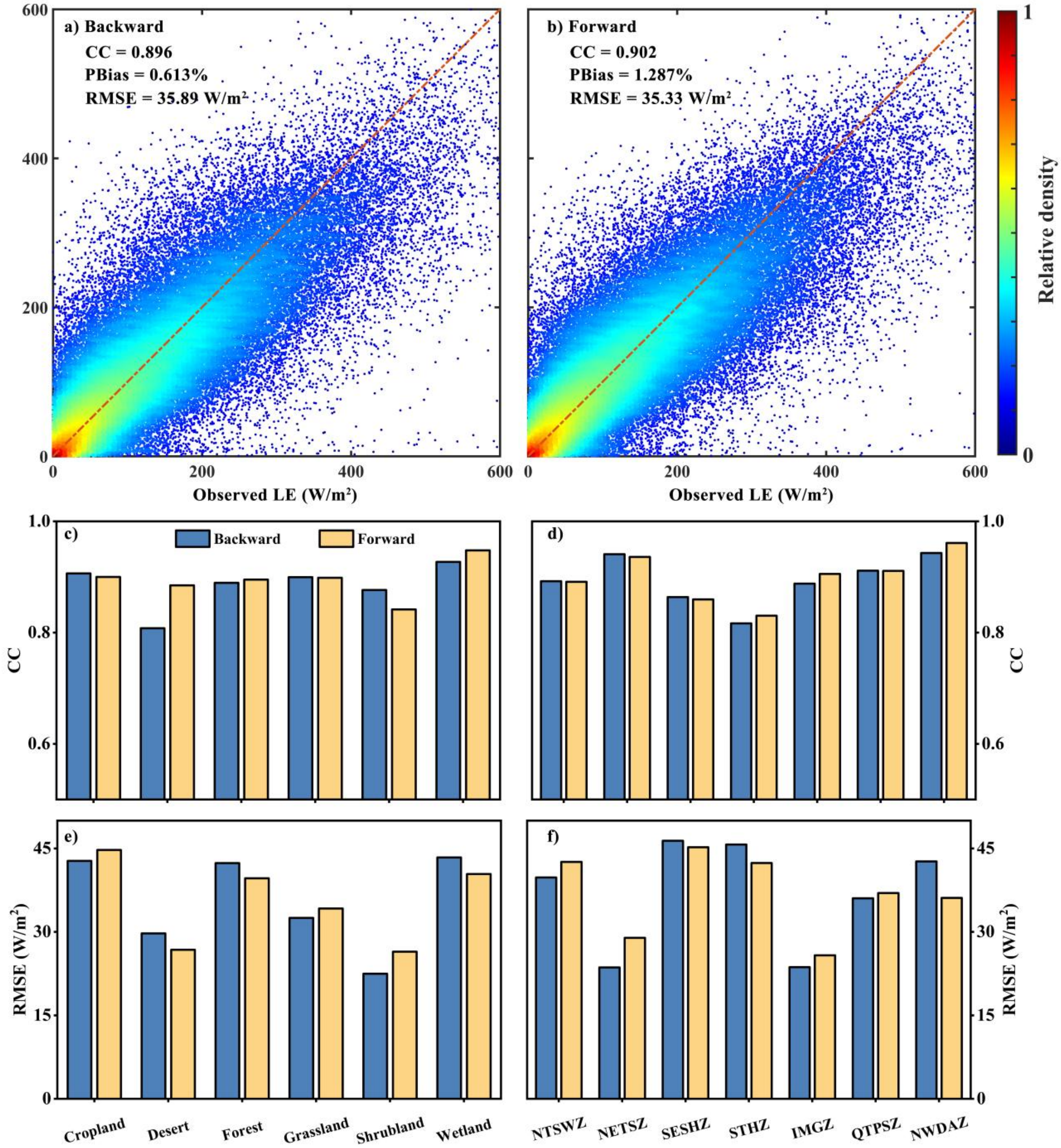


Figure 7: Consistency between backward (a) and forward (b) prolongation of half-hourly LE across underlying surface types (c,e) and climate zones (d,f).

3.2.2 Temporal stability of the prolongation

To evaluate the temporal stability of the prolongation results during long-term extrapolation, models were trained using either the first 2 years or the first 6 years of available observations, and their performance was subsequently assessed on a year-by-year basis for the following periods (Fig. 8). The 2-year training scenario represents an extreme case with minimal

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

training samples, whereas the 6-year training scenario is more representative of typical observation conditions at most sites. Overall, prolongation results under both training scenarios exhibit good temporal stability, with no evidence of systematic performance degradation as the extrapolation period increases. In terms of correlation, models trained on the first 2 years maintain relatively stable CC values in the range of approximately 0.88–0.93 across successive extrapolation years. Models trained on the first 6 years achieve slightly higher overall CC values and display reduced interannual variability, indicating that a larger training sample improves the consistency of prolongation results.

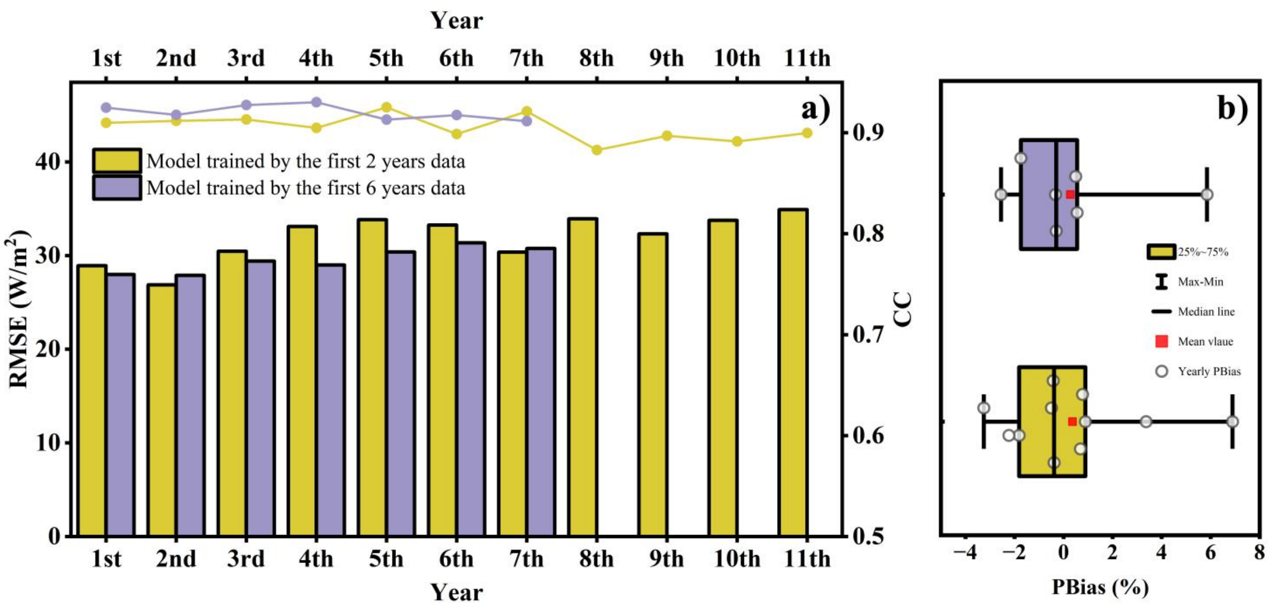


Figure 8: Temporal stability of forward prolongation performance for half-hourly LE using models trained with different lengths of observational data.

Error metrics show a similarly stable behavior. RMSE varies only modestly with extrapolation year, and models trained with 6 years of data generally produce lower RMSE values with smoother interannual variations. Bias analysis further indicates that PBias values under both training scenarios fluctuate around zero, with most years remaining within ±3 % and no systematic offset observed. In comparison, models trained on longer datasets exhibit more concentrated bias distributions. Taken together, these results demonstrate that the proposed temporal prolongation approach maintains high stability during long-term extrapolation. Even under the extreme scenario with limited training data, the method preserves reasonable correlations and error levels. As the length of the training dataset increases, interannual variability in CC, RMSE, and PBias is further reduced, confirming the suitability of the framework for constructing long-term continuous half-hourly LE time series.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

3.3 Demonstration of different scale prolonged time series

3.3.1 Half-hour scale

To illustrate the effects of temporal prolongation at the half-hourly scale, Fig. 9 presents comparisons between observed half-hourly latent heat flux (LE) and AutoML-based prolonged results at several representative sites spanning different underlying surface types and climate zones. Overall, the prolonged time series reasonably reproduce the diurnal cycles and amplitude characteristics of the observed LE, exhibiting stable temporal continuity and physically consistent structures. At cropland and grassland sites, the prolonged results closely match the observations in terms of the timing of peak values, diurnal variation patterns, and overall magnitude, indicating that the model effectively captures radiation-driven diurnal processes. For forest and wetland sites, where LE amplitudes are larger and short-term fluctuations are more complex, the prolonged series still track the main observed variations well, with only minor deviations occurring at a limited number of extreme peaks. In contrast, at sparsely vegetated sites such as desert and shrubland, LE magnitudes are relatively small and temporal variability is more irregular. Under these conditions, the prolonged results tend to be more conservative in representing isolated extreme high values but remain consistent with observations in terms of overall trends and diurnal rhythms. Similar patterns are observed across different climate zones, particularly in arid and high-altitude regions, where uncertainty is generally higher. Nevertheless, no evident nonphysical jumps or structural distortions are found in the prolonged time series. Overall, the half-hourly time series examples in Fig. 9 demonstrate that the proposed prolongation framework is able to stably reproduce high-frequency LE variability and diurnal cycle structures at the half-hourly scale, thereby providing a reliable basis for subsequent aggregation to daily and monthly timescales.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

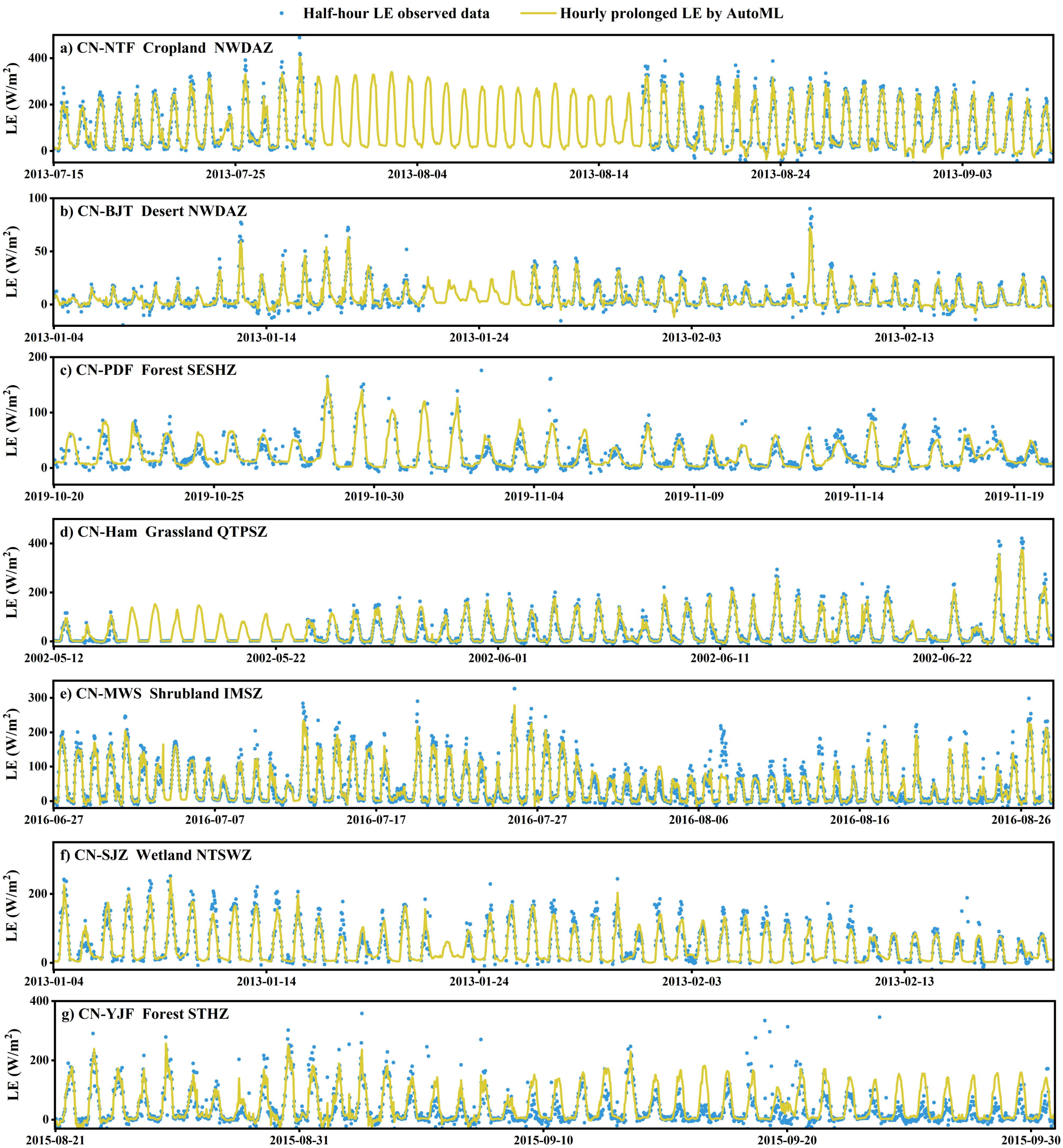


Figure 9: Demonstration of half-hourly prolonged LE time series across different typical sites.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

3.3.2 Daily scale

For the daily-scale analysis, only days with less than 10 % missing half-hourly LE data were considered as valid observations. Half-hourly LE records were aggregated to daily values and compared with the AutoML-based prolonged daily LE results. Figure 10 presents time series comparisons between prolonged daily LE and observation-based aggregated LE at representative sites. Overall, the prolonged daily LE results reasonably reproduce the seasonal variation patterns and intra-annual amplitude of the observed LE. Compared with the half-hourly scale, daily aggregation effectively smooths high-frequency noise, resulting in more continuous and stable temporal variations in the prolonged series. At cropland, grassland, and forest sites, the prolonged results show strong agreement with the observation-based daily LE in terms of peak timing, seasonal amplitude, and interannual variability, indicating that the model reliably captures evapotranspiration processes primarily controlled by radiation conditions and vegetation seasonality at the daily scale. At shrubland and wetland sites, daily LE variability is relatively higher, and some dispersion remains in the observed time series. Nevertheless, the prolonged results generally follow the main temporal patterns, with only minor deviations occurring during isolated high-value periods. Across different climate zones, the prolonged daily LE results do not exhibit evident systematic biases or trend-related errors, suggesting that the model maintains good temporal consistency and extrapolation stability when transitioning from the half-hourly to the daily scale. Overall, under conditions of limited valid observations, the prolonged daily LE data effectively preserve the seasonal structure and long-term variation of LE, providing a reliable data foundation for subsequent monthly aggregation and long-term change analyses.

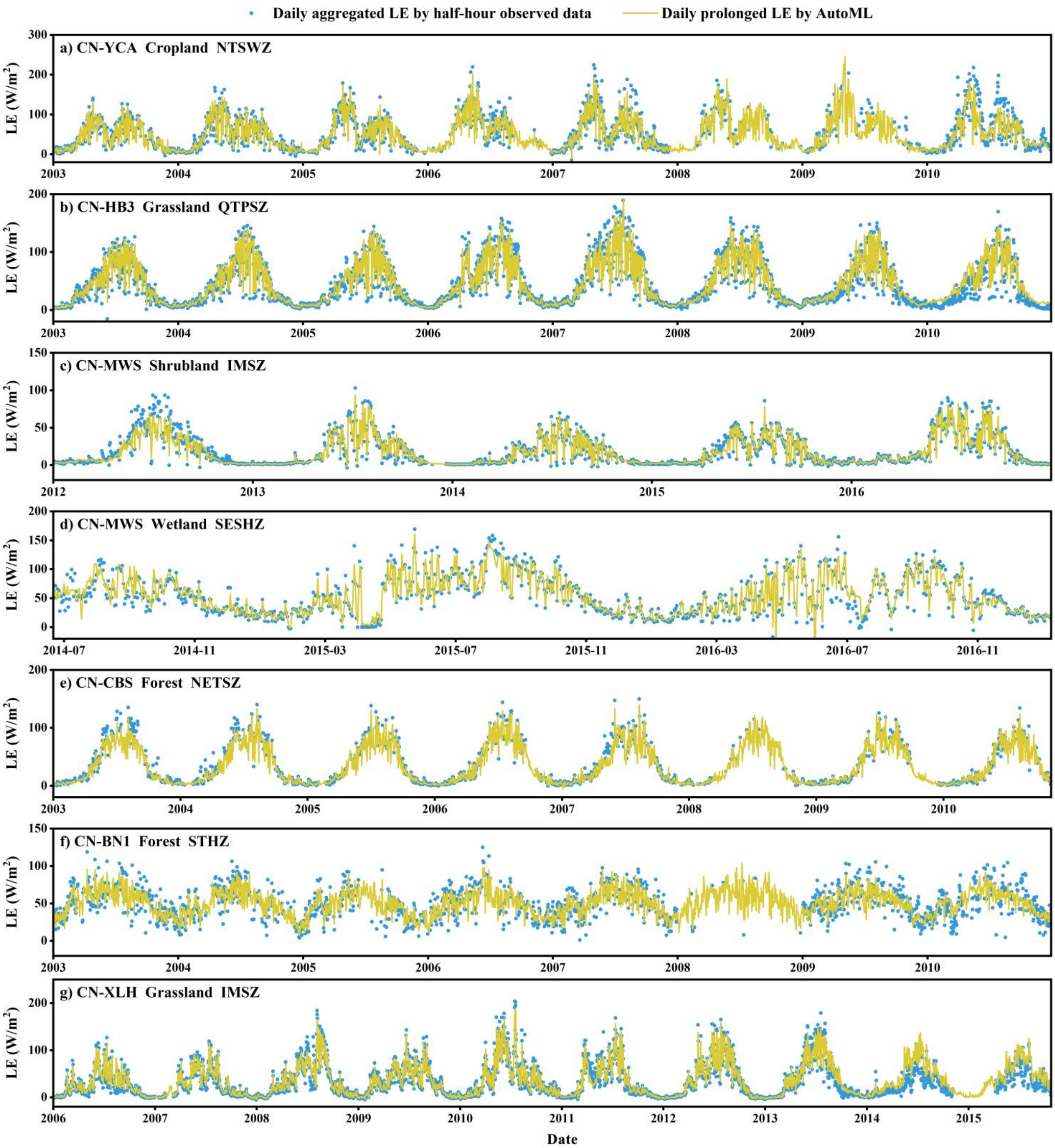
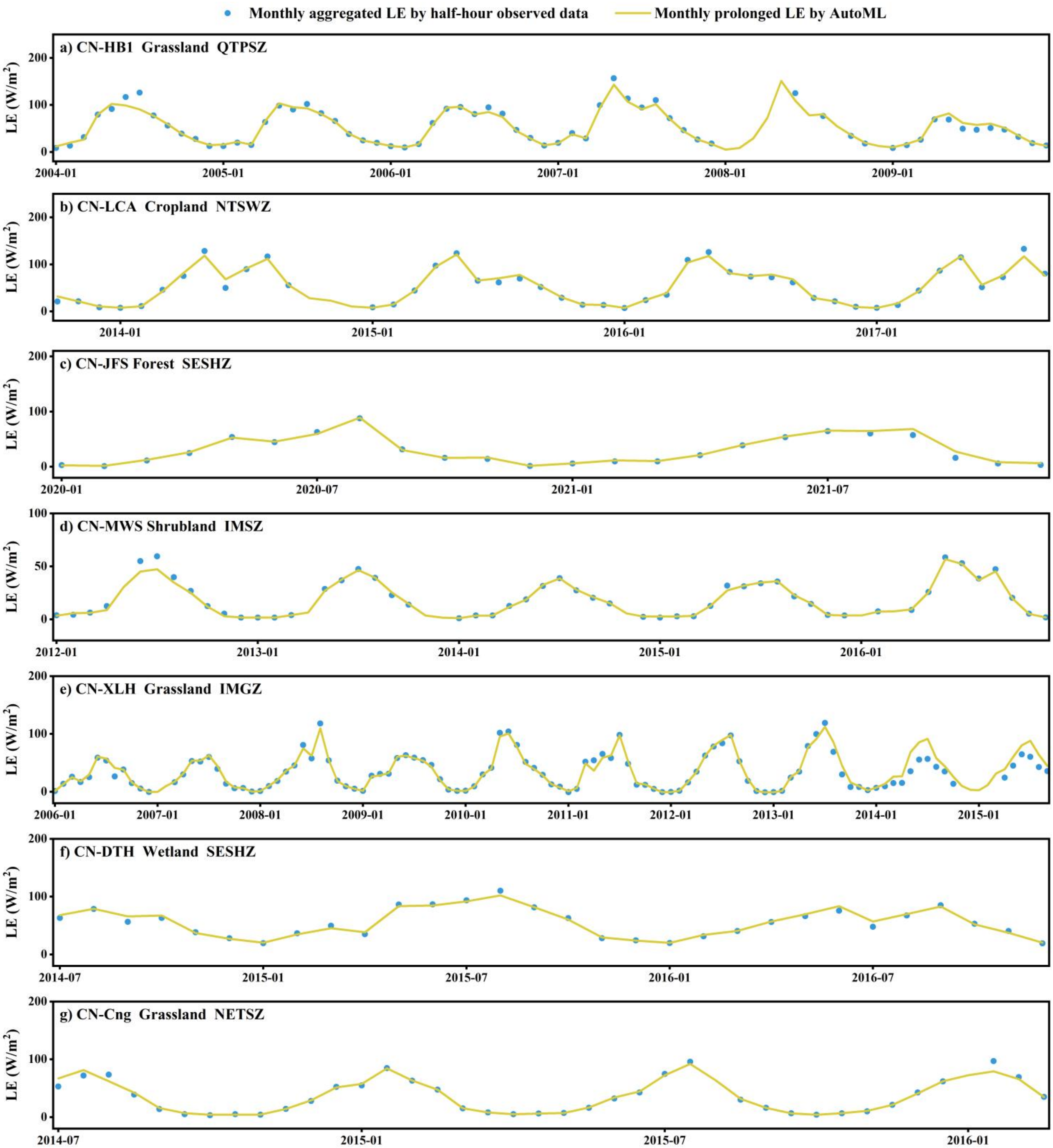


Figure 10: Demonstration of daily prolonged LE time series across different typical sites.

3.3.3 Monthly scale

For the monthly-scale analysis, only months with less than 10 % missing daily LE data were considered as valid observation samples. Daily LE observations were aggregated to monthly values and compared with the AutoML-based prolonged monthly LE results. Fig. 11 presents time series comparisons between prolonged monthly LE and observation-based aggregated LE at several representative sites. The prolonged monthly LE results stably reproduce the seasonal variation characteristics and interannual fluctuation trends of the observed LE. Compared with the half-hourly and daily scales, monthly aggregation further smooths short-term meteorological variability and high-frequency noise, resulting in more stable and continuous temporal behavior in the prolonged series. At grassland, cropland, and forest sites, the prolonged results closely match the observation-based monthly LE in terms of the timing of seasonal peaks, variation magnitude, and interannual trends, indicating that the model effectively captures evapotranspiration processes dominated by climatic background conditions and vegetation seasonality at the monthly scale. For underlying surface types such as shrubland and wetland, where observational samples are relatively limited and hydrothermal conditions are more complex, some uncertainty remains in monthly LE estimates. Nevertheless, the prolonged results generally follow the primary temporal patterns of the aggregated observations and do not exhibit evident systematic biases or trend-related errors. Across representative sites in different climate zones, the prolonged monthly LE data show good consistency, reasonably reflecting seasonal cycles and interannual variability of LE while maintaining temporal coherence across years. Overall, despite the relatively limited availability of monthly observation samples, the AutoML-based prolonged monthly LE results reliably reproduce the major seasonal structures and long-term trends of LE, providing stable data support for regional hydrological analyses, climate change studies, and the evaluation of multi-source evapotranspiration products.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026



510 **Figure 11: Demonstration of monthly prolonged LE time series across different typical sites.**

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

4 Discussion

4.1 Comparison between ChinaFlux and our dataset

The preceding sections have demonstrated the stability of the proposed gap-filling and temporal prolongation framework across multiple temporal scales. To further assess the reliability of the final dataset, we compared the constructed LE dataset with official ChinaFlux observations. Given the widespread occurrence of missing data in flux measurements, only observations with zero missing rates at the corresponding temporal scale were selected as reference data for comparison. As shown in Fig. 12, the two datasets exhibit high consistency in their statistical distributions at both the half-hourly and daily scales, considering the overall samples as well as stratifications by underlying surface type and climate zone. The medians and distribution ranges of the two datasets are comparable, indicating that the gap-filling and temporal prolongation procedures do not introduce evident systematic biases. This consistency is generally stable across most underlying surface types and climate zones, with relatively larger dispersion observed only in high-variability ecosystems such as desert and shrubland, yet without any directional bias.

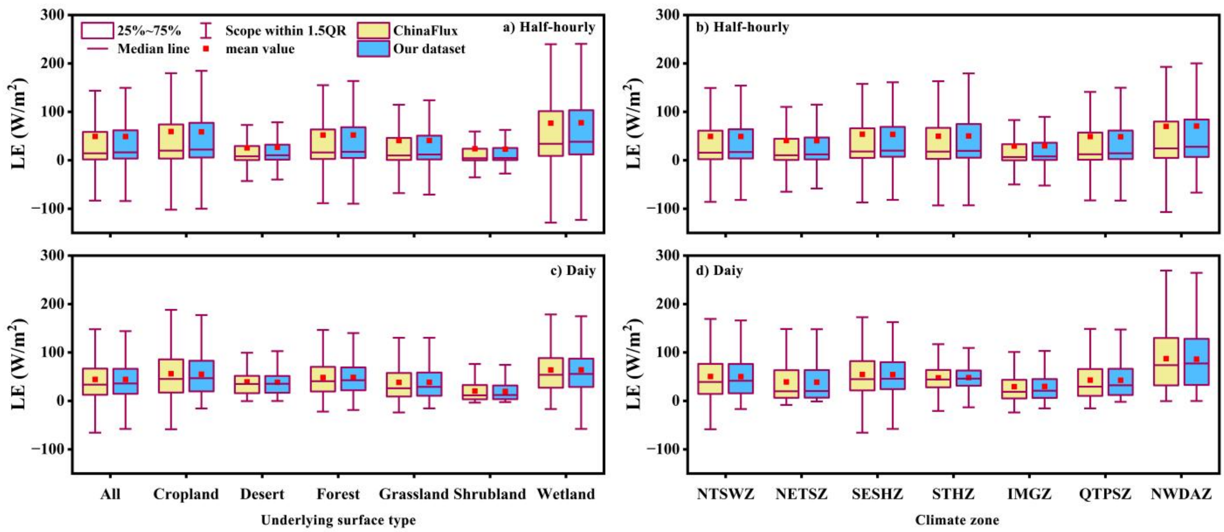


Figure 12: Comparison of data distributions between the reconstructed LE dataset and ChinaFlux observations at half-hourly (a,b) and daily (c,d) scales across different underlying surface types and climate zones.

Figure 13 presents comparisons aggregated by underlying surface type and climate zone at the daily, monthly, and annual scales. As the temporal scale progresses from daily to monthly and annual, the agreement between the two datasets further improves and dispersion decreases markedly, suggesting that the prolongation results effectively preserve the evapotranspiration characteristics reflected by ChinaFlux observations in a long-term mean sense. In particular, at the monthly and annual scales, the two datasets remain highly consistent in terms of variation magnitude and interannual trends, meeting the stability requirements for long-term hydrological and climate studies. It should be noted that the aggregated ChinaFlux results still exhibit some dispersion at specific sites and periods, primarily due to continuous data gaps within

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

certain months or years. In contrast, the dataset constructed in this study incorporates continuous reanalysis and remote sensing information to estimate LE at each time step during missing periods, thereby maintaining temporal continuity under conditions of limited observations. As a result, the constructed dataset exhibits more stable statistical characteristics at the monthly and annual scales. Overall, the LE dataset developed in this study shows good agreement with ChinaFlux observations across different temporal scales, underlying surface types, and climate zones. While effectively compensating for missing flux observations, it reasonably preserves the statistical distributions and long-term variability inherent in the original measurements, providing a reliable data foundation for regional water–energy cycle studies and the evaluation of evapotranspiration products.

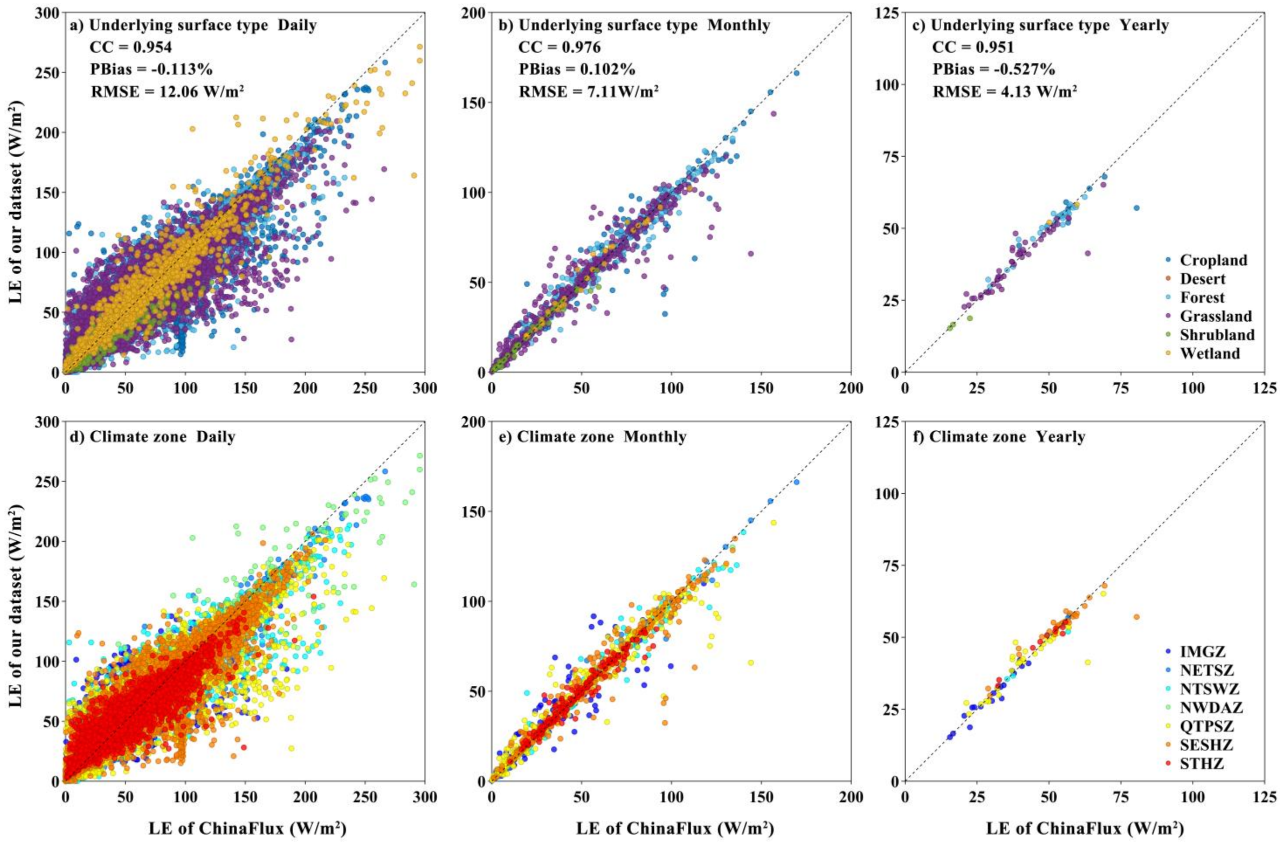


Figure 13: Comparison between the reconstructed LE dataset and ChinaFlux observations aggregated at daily (a,d), monthly (b,e), and yearly (c,f) scales for different underlying surface types and climate zones.

4.2 Importance of input features

To further understand how the model represents latent heat flux (LE) under different environmental conditions, the SHapley Additive exPlanations (SHAP) method was applied to quantify the relative importance of input features and their contributions to model predictions. Figures 14 and 15 present the SHAP results for the full sample and for stratifications by underlying surface type and climate zone, respectively, illustrating the relative contributions and distributions of individual

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

predictors to LE estimation. For the full sample (Fig. 14a), variables directly related to surface energy supply dominate the model, with net radiation (Rn) and energy-related terms associated with LE exhibiting the highest importance. This result indicates that the model broadly adheres to the fundamental physical constraints of evapotranspiration, whereby energy availability is the primary control on LE variability. At the same time, vegetation-related variables such as leaf area index (LAI) and normalized difference vegetation index (NDVI) also show substantial importance, highlighting the significant regulatory role of vegetation growth and seasonality in evapotranspiration processes. In contrast, variables characterizing atmospheric conditions and water availability (e.g., vapor pressure deficit (VPD), soil moisture, and precipitation) contribute less at the full-sample level, although they still play important complementary roles under specific environmental conditions. Across different underlying surface types (Fig. 14b–g), the ranking of feature importance exhibits pronounced differences. For cropland, forest, and grassland, energy-related variables and vegetation indices jointly dominate LE variability, indicating that in ecosystems with relatively dense vegetation cover, evapotranspiration is simultaneously controlled by radiative conditions and vegetation physiological activity. In contrast, for desert and shrubland, the importance of vegetation-related variables decreases markedly, while radiation- and atmosphere-related factors become relatively more influential, suggesting that under sparse vegetation conditions, LE is more directly constrained by energy input and atmospheric evaporative demand. Wetlands show relatively high contributions from both energy-related and water-related variables, reflecting the critical role of water availability in regulating evapotranspiration in these environments.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

Earth System Science Data
Open Access Discussions

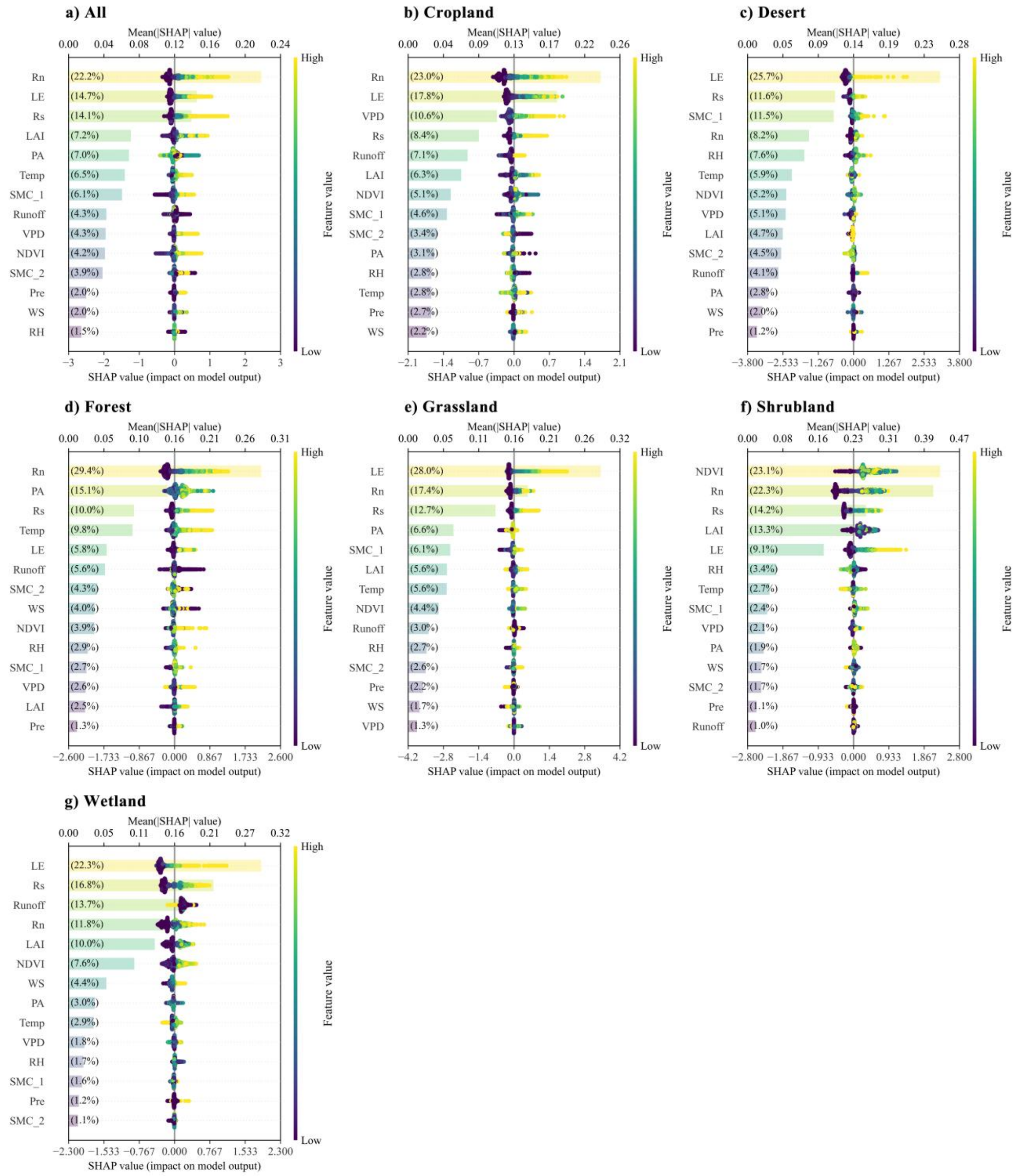


Figure 14: SHAP-based feature importance and contribution patterns for LE prediction across all samples and different underlying surface types.

The SHAP results across different climate zones (Fig. 15) further elucidate how the model responds to regional climatic differences. Overall, energy-related variables consistently dominate across all climate zones, but their relative rankings and secondary controlling factors vary substantially among regions. In humid and subhumid climate zones (e.g., NTSWZ,

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

NETSZ, and SESHZ), net radiation (Rn) and shortwave radiation (Rs) consistently rank among the most important predictors, indicating that LE variability is primarily controlled by surface energy supply. Meanwhile, LE itself and vegetation state variables such as LAI and NDVI contribute persistently, albeit to a lesser extent than energy terms, reflecting the modulation of evapotranspiration by vegetation seasonality. By contrast, atmospheric and soil moisture variables (e.g., VPD, soil moisture content (SMC), and precipitation) generally exhibit lower importance and mainly act as secondary regulators. In arid and semiarid climate zones (e.g., IMSZ, QTPSZ, and NWDAZ), the relative importance of energy-related variables (Rn and Rs) and LE further increases, while the contribution of vegetation indices declines markedly. At the same time, atmospheric variables such as air temperature and VPD enter the upper ranks of importance in some regions, indicating that under water-limited conditions, LE is more directly constrained by the combined effects of energy input and atmospheric evaporative demand. In particular, in the plateau climate zone (QTPSZ), air temperature and soil moisture variables also exhibit non-negligible importance in addition to radiative factors, reflecting the integrated modulation of evapotranspiration by complex terrain and low-temperature environments.

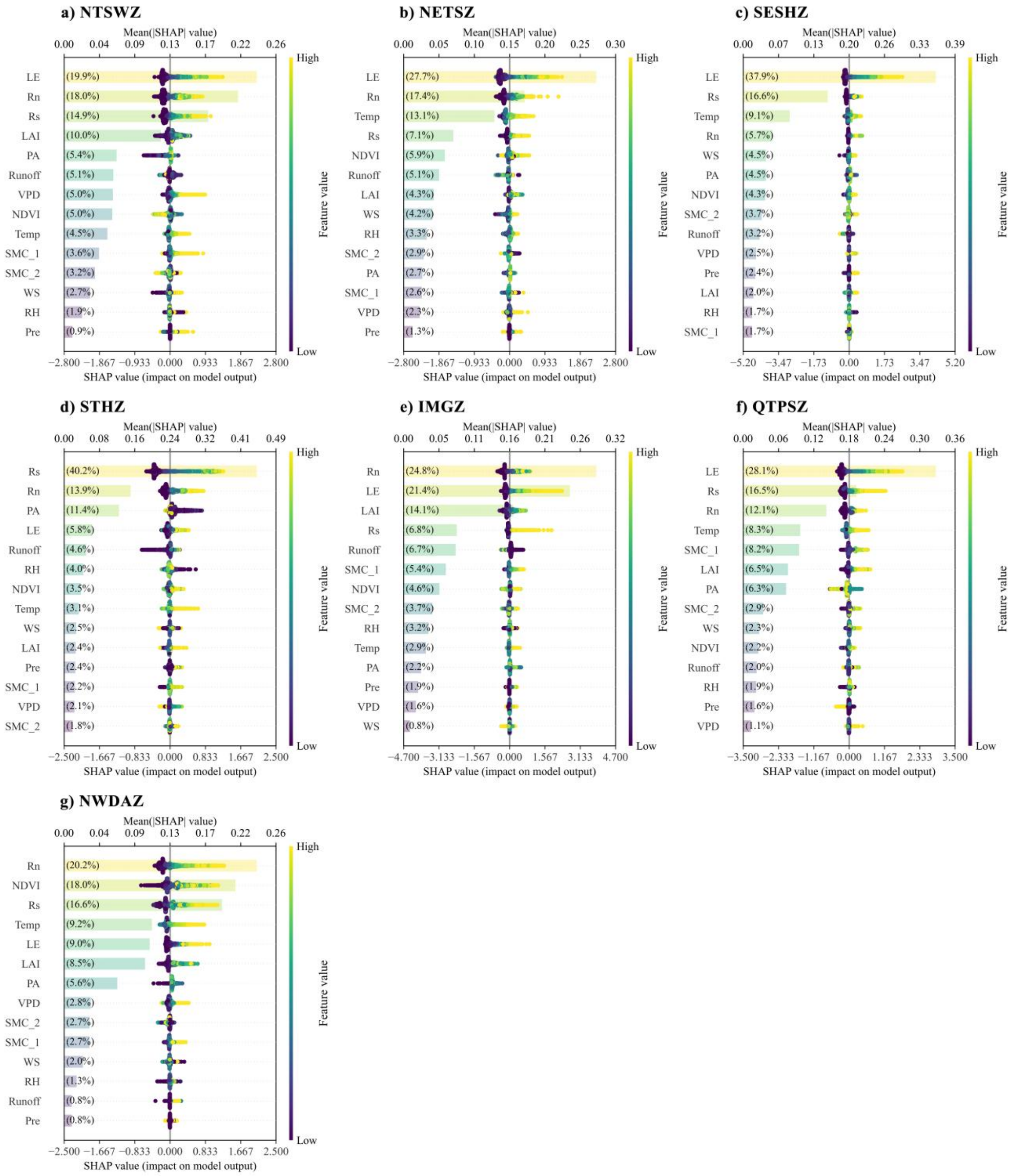


Figure 15: SHAP-based feature importance and contribution patterns for LE prediction across different climate zones.

585 ### 4.3 Contributions, Limitations, and Future Prospects

Based on a systematic evaluation of multi-source flux tower data, this study constructed a long-term, temporally continuous half-hourly latent heat flux (LE) dataset covering China and verified its reliability across multiple temporal scales. Compared

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

with conventional sliding-window-based gap-filling approaches, the AutoML framework adopted here demonstrates greater robustness under long continuous gap conditions and enables temporal prolongation of flux time series, providing a practical solution for generating continuous half-hourly LE records spanning 2000–2024. By integrating observations from more than 50 ChinaFlux sites, the resulting dataset substantially improves spatial representativeness and temporal continuity relative to previous studies, offering an important ground-based benchmark for the evaluation of remote sensing evapotranspiration products, land surface model validation, and regional water–energy cycle research.

Nevertheless, several limitations remain. First, model performance still varies across land cover types and climate zones, and uncertainties persist in representing extreme high LE values, particularly in sparsely vegetated or water-limited environments. Second, SHAP analyses indicate a strong model dependence on energy-related variables and vegetation indices. Although bias correction was applied to reanalysis and remote sensing inputs, uncertainties inherent in these drivers may still propagate into the final LE estimates. In addition, the temporal prolongation approach assumes that historical statistical relationships remain relatively stable over time, an assumption that warrants further examination under ongoing climate change.

Future work can be extended in several directions. On the one hand, the incorporation of higher-quality or higher spatiotemporal-resolution remote sensing and reanalysis products could further reduce uncertainties in key driving variables. On the other hand, integrating physical constraints or process-based information may enable the development of hybrid physics–data frameworks for flux gap-filling and temporal prolongation. Furthermore, as additional flux tower observations become available, the proposed approach and dataset could be extended to larger spatial domains, providing a more robust foundation for long-term evapotranspiration change analysis and multi-source data integration.

5 Data availability

This study publicly releases a temporally continuous half-hourly latent heat flux (LE) dataset for flux tower sites across China, together with multi-temporal-scale aggregated products and key driving variables used for gap-filling and temporal prolongation. The dataset covers the period from 2000 to 2024 and includes the following data products.

1) Half-hourly LE data (2000–2024). The half-hourly LE dataset represents the core product of this study and provides complete time series for all flux tower sites. Each record is accompanied by a quality flag indicating its data source: T denotes half-hourly LE derived from original flux tower observations; F represents gap-filled LE generated within the observation period using the AutoML-H2O framework; and P indicates LE values produced through temporal prolongation beyond the observation period using the same framework. All half-hourly data follow a unified time format, with timestamps referenced to local site time to ensure temporal continuity and traceability. Timestamps are provided in the format *YYYYMMDD–HHMM* (e.g., 20120101–1530), representing local date and time at each site..

2) Daily, monthly, and annual aggregated data. Daily, monthly, and annual LE products were directly aggregated from the gap-filled and temporally prolonged half-hourly LE time series and cover the period 2000–2024. All aggregated products

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

originate from the same half-hourly dataset, ensuring full temporal consistency across different time scales. For each temporal scale, quality control information is provided in the form of the proportion of valid half-hourly observations contributing to the aggregated value within the corresponding period (day, month, or year). This unified definition of data availability enables users to consistently assess the reliability of LE estimates across temporal scales while retaining stable and continuous datasets for long-term hydrological and climate change studies.

3) Auxiliary driving data. To enhance reproducibility and extensibility, key driving variables used for LE gap-filling and temporal prolongation are also provided. These include hourly meteorological and energy balance–related variables from ERA5-Land, as well as NDVI and LAI products from MODIS. All auxiliary datasets have been spatially and temporally matched to the LE data and can be directly used for method reproduction or extended analyses.

In addition, the code used for LE gap-filling and temporal prolongation is publicly available and provided alongside the dataset, enabling full reproducibility of the data processing workflow and facilitating reuse for similar applications.

All data products are released in CSV format, with file names clearly indicating temporal scale and variable type to facilitate data access and use. The product has been deposited at https://doi.org/10.5281/zenodo.18194590 (Qian et al., 2026) and can be downloaded publicly.

6 Conclusion

With the growing demand for long-term, temporally continuous ground-based benchmark data in climate change research, evapotranspiration modeling, and multi-source product evaluation, existing ChinaFlux observations remain limited by data discontinuity and substantial data gaps. To address these challenges, this study developed a gap-filling and temporal prolongation framework for half-hourly latent heat flux (LE) and, based on observations from more than 50 ChinaFlux sites, generated a continuous half-hourly LE dataset covering China for the period 2000–2024. The main conclusions are summarized as follows.

1) Gap-filling performance at the half-hourly scale. Across different gap-length scenarios, the AutoML approach consistently outperforms the traditional marginal distribution sampling (MDS) method and commonly used machine learning algorithms, exhibiting stable performance in terms of correlation and error control (overall CC = 0.862; RMSE = 33.75 W m^{-2}). As gap length increases, the performance degradation of AutoML is substantially smaller than that of MDS, with RMSE reductions of approximately 20 % and 25 % under the 7 d and 30 d gap scenarios, respectively. These advantages are particularly pronounced over high-uncertainty land cover types such as desert and shrubland.

2) Consistency and stability of temporal prolongation. Forward and backward temporal prolongation results show high consistency (CC = 0.902 and 0.896; RMSE = 35.33 and 35.89 W m^{-2}, respectively), indicating limited sensitivity to the direction of extrapolation. Temporal stability analyses further demonstrate that even when models are trained using only the first 2 years of observations, prolongation results maintain stable accuracy. With training samples extended to 6 years, interannual variability is further reduced, reflecting robust long-term stability.

3) Multi-temporal-scale consistency. From the half-hourly to daily and monthly scales, the prolonged LE data reasonably reproduce the main characteristics of observed LE variability. At the half-hourly scale, diurnal cycle structures are preserved; at the daily scale, seasonal variations are stably captured; and at the monthly scale, the prolonged data show strong agreement with observations in terms of seasonal cycles and interannual trends, meeting the requirements of long-term hydrological and climate studies.

4) Consistency and physical interpretability. Across different temporal scales, land cover types, and climate zones, the constructed dataset shows good agreement with ChinaFlux observations. SHAP analyses indicate that the model identifies physically meaningful controls on LE, with energy-related variables consistently dominating, while vegetation state and water availability modulate their relative importance under different environmental conditions. This enhances the interpretability and credibility of the dataset.

Overall, the proposed gap-filling and temporal prolongation framework enables reliable reconstruction of long-term continuous half-hourly LE time series across China, resulting in a ground-based benchmark dataset characterized by strong consistency, high stability, and clear physical constraints. This dataset provides valuable data support for the evaluation of remote sensing evapotranspiration products, land surface model validation, and regional water–energy cycle and climate change research.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

Earth System Science Data Discussions
Open Access

Appendix A

Table A1. Basic information for 50 sites in China (The * in the station ID indicates that this station provides only interpolated data.).

Station ID	SITE Name	Longitude (°E)	Latitude (°N)	Climate zone	Underlying surface type	Elevation (m)	Annual average temperature (°C)	Annual precipitation (mm)	Number of LE Data	Missing ratio	Start year	End year
1*	CN-JFS	107.1508	29.0217	SESHZ	Forest	2	9.5	631	35088	0.000	2020	2021
2*	CN-NQF	92.0167	31.6500	QTPSZ	Grassland	1409	4.6	255	87648	0.000	2014	2018
3*	CN-QYZ	115.0667	26.7333	SESHZ	Forest	111	17.9	1489	227904	0.000	2003	2015
4*	CN-XLD	112.4667	35.0167	NTSWZ	Forest	3500	1.5	747	35088	0.000	2016	2017
5*	CN-DHS	112.5343	23.1738	STHZ	Forest	1317	2.1	365	140256	0.000	2003	2011
6*	CN-HB3	101.3333	37.6667	QTPSZ	Grassland	145	5.6	420	140256	0.000	2003	2011
7*	CN-DXF	91.0833	30.8500	QTPSZ	Grassland	1411	15.2	830	122736	0.000	2004	2011
8*	CN-NMG	116.4040	43.3255	IMGZ	Grassland	1350	2.4	375	122736	0.000	2004	2011
9	CN-DSL	98.9406	38.8399	QTPSZ	Wetland	553	23.8	712	44427	0.000	2014	2016
10*	CN-YJF	102.1775	23.4739	STHZ	Grassland	4333	1.3	477	46800	0.001	2013	2016
11*	CN-BTM	111.9352	33.4997	SESHZ	Forest	1378	9.0	124	35040	0.001	2017	2019
12	CN-CLF	123.4703	44.5966	NETSZ	Cropland	3200	-1.7	580	43617	0.009	2018	2021
13	CN-Cng	123.5092	44.5934	NETSZ	Grassland	2400	21.5	1931	58312	0.023	2007	2010
14	CN-HaM	101.1800	37.3700	QTPSZ	Grassland	1080	3.8	800	50439	0.036	2002	2004
15	CN-MWS	107.2300	37.7100	IMSZ	Shrubland	15	12.2	12	83863	0.044	2012	2016
16	CN-HB1	101.3167	37.6167	QTPSZ	Grassland	143	6.4	500	100449	0.045	2004	2009
17	CN-LCA	114.6833	37.8833	NTSWZ	Cropland	876	9.8	37	66910	0.046	2013	2017
18	CN-XLH	116.6714	43.5544	IMGZ	Grassland	1350	2.4	380	162766	0.053	2006	2015
19	CN-JRF	119.2173	31.8068	SESHZ	Cropland	874	9.8	37	98580	0.059	2015	2020
20	CN-GCF	115.6667	39.1333	NTSWZ	Cropland	3480	1.2	720	43351	0.136	2020	2022
21	CN-YCA	116.5702	36.8290	NTSWZ	Cropland	3680	-1.8	580	120296	0.143	2003	2011
22	CN-REG	102.5500	32.8000	QTPSZ	Grassland	1556	7.3	185	83600	0.153	2015	2020
23	CN-DTH	113.0525	29.4875	SESHZ	Wetland	3850	-2.4	420	39638	0.197	2006	2008
24	CN-BN1	101.2653	21.9275	STHZ	Forest	1280	8.8	510	108316	0.228	2003	2011
25	CN-CBS	128.0958	42.4025	NETSZ	Forest	1530	8.3	293	99328	0.278	2003	2010
26	CN-JZF	121.2017	41.1480	NTSWZ	Cropland	1054	9.8	37	110654	0.369	2005	2014
27	CN-PJS	121.9646	40.9328	NTSWZ	Wetland	181	12.5	580	31866	0.393	2018	2020
28	CN-SJZ	118.9809	37.7664	NTSWZ	Wetland	328	12.5	580	84696	0.396	2011	2018
29	CN-Du2	116.2836	42.0466	IMGZ	Grassland	592	21.0	1490	39647	0.413	2015	2018
30	CN-HZF	121.0178	51.7811	NETSZ	Forest	1170	16.0	1432	49790	0.432	2014	2018
31	CN-Hgu	102.5900	32.8453	QTPSZ	Grassland	4520	-2.1	480	25881	0.438	2015	2017

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

32	CN-PDF	106.3167	26.6000	SESHZ	Forest	18	15.8	1090	46063	0.451	2015	2019
33	CN-LDF	101.1326	41.9993	NWDAZ	Desert	210	13.4	642	22885	0.473	2013	2015
34	CN-YS1	116.6563	40.4190	NTSWZ	Shrubland	3150	0.8	520	18426	0.473	2020	2021
35	CN-SJY	100.4800	34.3547	QTPSZ	Grassland	3439	-1.2	360	45943	0.476	2012	2016
36	CN-SSW	100.4933	38.7892	NWDAZ	Desert	1250	2.6	349	20520	0.486	2013	2015
37	CN-HB2	101.3119	37.6094	QTPSZ	Grassland	4	15.3	1022	53371	0.493	2015	2020
38	CN-NTF	101.1338	42.0048	NWDAZ	Cropland	1731	7.3	185	20381	0.493	2013	2015
39	CN-HYL	101.1239	41.9932	NWDAZ	Forest	875	9.8	37	21210	0.507	2013	2015
40	CN-ARF	100.4643	38.0473	QTPSZ	Forest	50	13.4	341	59548	0.515	2013	2019
41	CN-YS3	116.6588	40.4165	NTSWZ	Forest	3950	-2.9	531	16344	0.534	2020	2021
42	CN-ZYF	100.4464	38.9751	NWDAZ	Wetland	1460	7.3	185	75913	0.567	2013	2022
43	CN-BJT	100.3042	38.9150	NWDAZ	Desert	4	12.7	604	14415	0.571	2013	2014
44	CN-DMF	110.3315	41.6439	IMGZ	Grassland	756	21.5	1557	69214	0.605	2013	2022
45	CN-HZZ	100.3186	38.7652	NWDAZ	Desert	1502	3.8	280	66672	0.620	2013	2022
46	CN-YKF	100.2421	38.0142	QTPSZ	Grassland	144	25.0	1607	26064	0.628	2015	2018
47	CN-HHL	101.1335	41.9903	NWDAZ	Forest	4585	1.9	430	58114	0.647	2013	2022
48	CN-DMZ	100.3722	38.8555	NWDAZ	Cropland	3200	-1.7	550	38125	0.660	2013	2019
49	CN-HMF	100.9872	42.1135	NWDAZ	Desert	1594	7.3	185	34189	0.746	2015	2022
50	CN-QJF	118.2500	31.9667	SESHZ	Cropland	15	15.7	1080	12543	0.762	2017	2020

Author contributions

Conceptualization: Long Qian and Lifeng Wu. Methodology: Long Qian, Lifeng Wu, and Xingjiao Yu. Data curation: Long Qian. Funding acquisition: Lifeng Wu, Zhitao Zhang, and Junying Chen. Writing (initial): Long Qian. Writing (review and editing): Xiaogang Liu, Sien Li and Sumeng Ye. Supervision: Rangjian Qiu Xuqian Bai, and Yaokui Cui.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Disclaimer

Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements

The authors gratefully acknowledge the ChinaFlux research network for providing eddy covariance flux observations used in this study. The ChinaFlux data were obtained through the National Ecosystem Science Data Center (https://www.nesdc.org.cn/, last access: 11 November 2025). We thank the European Centre for Medium-Range Weather Forecasts (ECMWF) for making the ERA5-Land reanalysis data publicly available (https://www.ecmwf.int/en/era5-land, last access: 11 November 2025). We also acknowledge the MODIS science team for providing the MOD13Q1, MYD13Q1, MOD15A2H, and MYD15A2H products used to derive vegetation indices in this study. The authors thank the Google Earth Engine platform (https://earthengine.google.com/, last access: 11 November 2025) for enabling efficient access to and processing of the ERA5-Land and MODIS datasets. We further acknowledge the H2O.ai team for developing the H2O AutoML framework (https://www.h2o.ai/, last access: 11 November 2025), which was used for gap-filling and temporal prolongation of latent heat flux time series.

Financial support

This research has been supported by the Special project of scientific and technological innovation of Xinjiang Research Institute of Arid Area Agriculture (XJHQNY-2025-3), Tianjin Key Laboratory of Rail Transit Navigation Positioning and Spatio-temporal Big Data Technology under Grant (TKL2024B12), National Key Research and Development Program of China (2023YFD1900702-01, 2022YFD1900404), the National Natural Science Foundation of China (52479050).

References

Allen, R. G., Pereira, L. S., Howell, T. A., and Jensen, M. E.: Evapotranspiration information reporting: I. Factors governing measurement accuracy, Agric. Water Manag., 98, 899–920, https://doi.org/10.1016/j.agwat.2010.12.015, 2011.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration – guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper No. 56, FAO, Rome, Italy, 1998.

Bellvert, J., Pamies-Sans, M., Casadesus, J., and Girona, J.: Evaluating the impact of drought and water restrictions on agricultural production in irrigated areas through crop water productivity functions and a remote sensing-based evapotranspiration model, Agric. Water Manag., 309, 109319, https://doi.org/10.1016/j.agwat.2025.109319, 2025.

Chen, Y. Y., Chu, C. R., and Li, M. H.: A gap-filling model for eddy covariance latent heat flux: Estimating evapotranspiration of a subtropical seasonal evergreen broad-leaved forest as an example, J. Hydrol., 468–469, 101–110, https://doi.org/10.1016/j.jhydrol.2012.08.026, 2012.

Cheng, M., Jiao, X., Jin, X., Li, B., Liu, K., and Shi, L.: Satellite time series data reveal interannual and seasonal spatiotemporal evapotranspiration patterns in China in response to effect factors, Agric. Water Manag., 255, 107046, https://doi.org/10.1016/j.agwat.2021.107046, 2021.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

Chu, R. H., Li, M., Towfiqul Islam, A. R. M., Fei, D. Y., and Shen, S. H.: Attribution analysis of actual and potential evapotranspiration changes based on the complementary relationship theory in the Huai River basin of eastern China, Int. J. Climatol., 39, 4072–4090, https://doi.org/10.1002/joc.6060, 2019.

Derardja, B., Khadra, R., Abdelmoneim, El-Shirbeny, M. A., Valsamidis, T., Pasquale, V. D., Deflorio, A. M., and Volden, E.: Advancements in remote sensing for evapotranspiration estimation: A comprehensive review of temperature-based models, Remote Sens., 16, 1927, https://doi.org/10.3390/rs16111927, 2025.

Elnashar, A., Wang, L. L., Wu, B. F., Zhu, W. W., and Zeng, H. W.: Synthesis of global actual evapotranspiration from 1982 to 2019, Earth Syst. Sci. Data, 13, 447–480, https://doi.org/10.5194/essd-13-447-2021, 2021.

Fan, J. L., Yue, W. J., Wu, L. F., Zhang, F. C., Cai, H. J., Wang, X. K., Lu, X. H., and Xiang, Y. Z.: Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China, Agric. For. Meteorol., 263, 225–241, https://doi.org/10.1016/j.agrformet.2018.08.019, 2018.

Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M. F., Hook, S., Baldocchi, D., Townsend, P. A., Kilic, A., Tu, K., Miralles, D. D., Perret, J., Lagouarde, J. P., Waliser, D., Purdy, A. J., French, A., Schimel, D., Famiglietti, J. S., Stephens, G., and Wood, E. F.: The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources, Water Resour. Res., 53, 2618–2626, https://doi.org/10.1002/2016WR020175, 2017.

Fu, J., Gong, Y. Q., Zheng, W. W., Zou, J., Meng, Z., Zhang, Z. B., Qin, J. X., Liu, J. X., and Quan, B.: Spatial-temporal variations of terrestrial evapotranspiration across China from 2000 to 2019, Sci. Total Environ., 825, 153951, https://doi.org/10.1016/j.scitotenv.2022.153951, 2022.

Fu, T. L., Li, X. R., Jia, R. L., and Feng, L.: A novel integrated method based on a machine learning model for estimating evapotranspiration in dryland, J. Hydrol., 603, 126881, https://doi.org/10.1016/j.jhydrol.2021.126881, 2021.

Gebrechorkos, S. H., Sheffield, J., Vicente-Serrano, S. M., Funk, C., Miralles, D. G., Peng, J., Dyer, E., Talib, J., Beck, H. E., Singer, M. B., and Dadson, S.: Warming accelerates global drought severity, Nature, https://doi.org/10.1038/s41586-025-09047-2, 2025.

Gu, L., Schumacher, D. L., Fischer, E. M., Slater, L. J., Yin, J. B., Sippel, S., Chen, J., Liu, P., and Knutti, R.: Flash drought impacts on global ecosystems amplified by extreme heat, Nat. Geosci., https://doi.org/10.1038/s41561-025-01719-y, 2025.

Guo, D., Parehkar, A., Ryu, D., Wang, Q. J., and Western, A. W.: Parsimonious gap-filling models for sub-daily actual evapotranspiration observations from eddy-covariance systems, Remote Sens., 14, 1286, https://doi.org/10.3390/rs14051286, 2022.

Guo, L. N., Wu, Y. H., Zheng, H. X., Zhang, B., Fan, L. X., Chi, H. J., Yan, B. K., and Wang, X.Q.: Consistency and uncertainty of gridded terrestrial evapotranspiration estimations over China, J. Hydrol., 612, 128245, https://doi.org/10.1016/j.jhydrol.2022.128245, 2022.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

He, X. L., Liu, S. M., Bateni, S., Xu, T. R., Jun, C., Kim, D., Li, X., Song, L. S., Zhao, L., Xu, Z. W., and Wei, J. X.: Innovative approach for estimating evapotranspiration and gross primary productivity by integrating land data assimilation, machine learning, and multi-source observations, Agric. For. Meteorol., 355, 110136, https://doi.org/10.1016/j.agrformet.2024.110136, 2025.

Jin, L., Chen, S. D., Yang, H. B., and Zhang, C. C.: Evaluation and drivers of four evapotranspiration products in the Yellow River Basin, Remote Sens., 16, 1829, https://doi.org/10.3390/rs16111829, 2024.

Jung, M., Koirala, S., Weber, U., Ichii, K., and Reichstein, M.: The FLUXCOM ensemble of global land–atmosphere energy fluxes, Sci. Data, 6, 74, https://doi.org/10.1038/s41597-019-0076-8, 2019.

Ke, Z., Kimball, J. S., and Running, S. W.: A review of remote sensing-based actual evapotranspiration estimation, WIREs Water, 3, 834–853, https://doi.org/10.1002/wat2.1168, 2016.

Khan, M. S., Jeon, S. B., and Jeong, M. H.: Gap-filling eddy covariance latent heat flux: Inter-comparison of four machine learning model predictions and uncertainties in forest ecosystem, Remote Sens., 13, 4976, https://doi.org/10.3390/rs13244976, 2021.

Kim, Y., Johnson, M. S., Knox, S. H., Black, T. A., Dalmagro, H. J., Kang, M., Kim, J., and Baldocchi, D.: Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis, Global Change Biol., 26, 1499–1518, https://doi.org/10.1111/gcb.1484, 2020.

Li, C. Y., Yuan, X., Jiao, Y., Ji, P., and Huang, Z. W.: High-resolution land surface modeling of the irrigation effects on evapotranspiration over the Yellow River basin, J. Hydrol., 633, 130986, https://doi.org/10.1016/j.jhydrol.2024.130986, 2024.

Li, C. Y., Yuan, X., Jiao, Y., Ji, P., and Huang, Z. W.: High-resolution land surface modeling of the irrigation effects on evapotranspiration over the Yellow River basin, J. Hydrol., 633, 130986, https://doi.org/10.1016/j.jhydrol.2024.130986, 2025.

Li, S. J., Wang, G. J., Sun, S. L., Hagan, D. F. T., Chen, T. X., Dolman, H., and Liu, Y.: Long-term changes in evapotranspiration over China and attribution to climatic drivers during 1980–2010, J. Hydrol., 595, 126037, https://doi.org/10.1016/j.jhydrol.2021.126037, 2021.

Li, W. Y. P., Yao, Z. Y., Qu, Y. F., Yang, H. B., Song, Y., Song, L. S., Wu, L. F., and Cui, Y. K.: A benchmark dataset for global evapotranspiration estimation based on FLUXNET2015 from 2000 to 2022, Earth Syst. Sci. Data, https://doi.org/10.5194/essd-2024-460, 2024.

Li, X., Pang, Z. J., Xue, F. H., Ding, J. L., Wang, J. J., Xu, T. R., Xu, Z. W., Ma, Y. F., Zhang, Y., and Shi, J. L.: Analysis of spatial and temporal variations in evapotranspiration and its driving factors based on multi-source remote sensing data: A case study of the Heihe River Basin, Remote Sens., 16, 2696, https://doi.org/10.3390/rs16152696, 2024.

Li, Y., Liu, X. N., Zhang, X. G., Gu, X. B., Yu, L. Y., Cai, H. J., and Peng, X. B.: Using solar-induced chlorophyll fluorescence to predict winter wheat actual evapotranspiration through machine learning and deep learning methods, Agric. Water Manag., 309, 109322, https://doi.org/10.1016/j.agwat.2025.109322, 2025.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

Lian, X., Piao, S., and Huntingford, C., Li, Y., Zeng, Z. Z., Wang, X. H., Ciais, P., McVicar, T. R., Peng, S. S., Ottlé, C., Yang, H., Yang, Y. T., Zhang, Y. Q., and Wang, T.: Partitioning global land evapotranspiration using CMIP5 models constrained by observations, Nat. Clim. Change, 8, 640–646, https://doi.org/10.1038/s41558-018-0207-9, 2018.

Liu, Y., Zhang, S., Zhang, J. H., Tang, L. L., and Bai, Y.: Assessment and comparison of six machine learning models in estimating evapotranspiration over croplands using remote sensing and meteorological factors, Remote Sens., 13, 3838, https://doi.org/10.3390/rs13193838, 2021.

Long, D., Longuevergne, L., and Scanlon, B. R.: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites, Water Resour. Res., 50, https://doi.org/10.1002/2013WR014581, 2014.

Lu, J., Wang, G. J., Chen, T. X., Li, S. J., Hagan, D. F. T., Kattel, G., Peng, J., and Jiang, T.: A harmonized global land evaporation dataset from model-based products covering 1980–2017, Earth Syst. Sci. Data, 13, 5879–5898, https://doi.org/10.5194/essd-13-5879-2021, 2021.

Ma, N., Szilagyi, J., and Zhang, Y.: Calibration-free complementary relationship estimates terrestrial evapotranspiration globally, Water Resour. Res., 57, e2021WR029691, https://doi.org/10.1029/2021WR029691, 2021.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115, 1781–1800, https://doi.org/10.1016/j.rse.2011.02.019, 2011.

Mu, Q. Z., Zhao, M. S., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115, 1781–1800, https://doi.org/10.1016/j.rse.2011.02.019, 2011.

Nourani, V., Ahmadi, R., Zhang, Y. Q., and Dąbrowska, D.: Ensemble machine learning-based extrapolation of Penman–Monteith–Leuning evapotranspiration data, Ecol. Indic., 170, 113012, https://doi.org/10.1016/j.ecolind.2024.113012, 2025.

Oki, T., and Kanae, S.: Global hydrological cycles and world water resources, Science, 313, 1068–1072, https://doi.org/10.1126/science.1128845, 2006.

Pascolini-Campbell, M., Reager, J. T., Chandanpurkar, H. A., and Rodell, M.: A 10 per cent increase in global land evapotranspiration from 2003 to 2019, Nature, 593, 543–547, https://doi.org/10.1038/s41586-021-03503-5, 2021.

Qian, L., Zhang, Z. T., Wu, L. F., Fan, S. S., Yu, X. J., Liu, X. G., Ba, Y. L., Ma, H. J., Wang, Y.C.: High uncertainty of evapotranspiration products under extreme climatic conditions. J. Hydrol., 626, 130332, https://doi.org/10.1016/j.jhydrol.2023.130332, 2023.

Qian, L., Yu, X. J., Wu, L. F., Zhang, Z. T., Fan, S. L., Du, R. Q., Liu, X. G., Yang, Q. L., Qiu, R. J., Cui, Y. K., Huang, G. M., and Wang, Y. C.: Improving high uncertainty of evapotranspiration products under extreme climatic conditions based on deep learning and ERA5 reanalysis data, J. Hydrol., 641, 131755, https://doi.org/10.1016/j.jhydrol.2024.131755, 2024a.

Qian, L., Wu, L. F., Zhang, Z. T., Fan, J. L., Yu, X. J., Liu, X. G., Yang, Q. L., Cui, Y. K.: A gap-filling method for daily evapotranspiration of global flux data sets based on deep learning, J. Hydrol., 641, 131787, https://doi.org/10.1016/j.jhydrol.2024.131787, 2024b.

https://doi.org/10.5194/essd-2026-22
Preprint. Discussion started: 12 February 2026

Qian, L., Yu, X. J., Wu, L. F., Cui, Y. K., Zhang, Z. T., Chen, J. Y., Ye, S. M., Bai, X. Q., Liu, X. G., Li, S. E., and Qiu, R. J.: A benchmark dataset for half-hourly evapotranspiration estimation in China from 2000 to 2024 (Version v1). Zenodo [data set], https://doi.org/10.5281/zenodo.18194590, 2026.

Rodell, M., Beaudoing, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayson, C. A., Chambers, D., Clark, E., Fetzer, E. J., Gao, X., Gu, G., Hilburn, K., Huffman, G. J., Lettenmaier, D. P., Liu, W. T., Robertson, F. R., Schlosser, C. A., Sheffield, J., and Wood, E. F.: The observed state of the water cycle in the early twenty-first century, J. Clim., 28, 8289–8318, https://doi.org/10.1175/JCLI-D-14-00555.1, 2015.

Samaniego, L.: Permanent shifts in the global water cycle, Science, 387, 6741, https://doi.org/10.1126/science.adw5851, 2025.

Tang, R. L., Peng, Z., Liu, M., Li, Z. L., Jiang, Y. Z., Hu, Y. X., Huang, L. X., Wang, Y. Z., Wang, J. R., Jia, L., Zheng, C. L., Zhang, Y. Q., Zhang, K., Yao, Y. J., Chen, X. L., Xiong, Y. J., Zeng, Z. Z., and Fisher, J. B.: Spatial-temporal patterns of land surface evapotranspiration from global products, Remote Sens. Environ., 304, 114066, https://doi.org/10.1016/j.rse.2024.114066, 2024.

Trenberth, K. E., Fasullo, J. T., and Kiehl, J.: Earth's global energy budget, Bull. Am. Meteorol. Soc., 90, 311–323, https://doi.org/10.1175/2008BAMS2634.1, 2009.

Wang, Y. R., Hessen, D. O., Samset, B. H., and Stordal, F.: Evaluating global and regional land warming trends in the past decades with both MODIS and ERA5-Land land surface temperature data, Remote Sens. Environ., 280, 113181, https://doi.org/10.1016/j.rse.2022.113181, 2022.

Wu, T. A., Zhang, W., Jiao, X. Y., Guo, W. H., and Hamoud, Y. A..: Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration, Comput. Electron. Agric., 184, 106039, https://doi.org/10.1016/j.compag.2021.106039, 2021.

Wu, X. Y., Han, W. Y., and Yu, Z.: Evapotranspiration increment was underestimated in China due to underrepresented land cover changes, Environ. Res. Lett., 19, 064063, https://doi.org/10.1088/1748-9326/ad4977, 2024.

Xiang, K. Y., Li, Y., Horton, R., and Feng, H.: Similarity and difference of potential evapotranspiration and reference crop evapotranspiration – a review, Agric. Water Manag., 232, 106043, https://doi.org/10.1016/j.agwat.2020.106043, 2020.

Xu, C. Y., and Singh, V. P.: Evaluation of three complementary relationship evapotranspiration models by water balance approach to estimate actual regional evapotranspiration in different climatic regions, J. Hydrol., 308, 105–121, https://doi.org/10.1016/j.jhydrol.2004.10.024, 2005.

Xu, Q. C., Li, L., Wei, Z. W., et al.: A multimodal machine learning fused global 0.1° daily evapotranspiration dataset from 1950–2022, Agric. For. Meteorol., 372, 110645, https://doi.org/10.1016/j.agrformet.2025.110645, 2025.

Yang, D. N., Yang, S. S., Huang, J. J., Zhang, S. Y., Zhang, S., Zhang, J. H., and Bai, Y.: Improving time upscaling of instantaneous evapotranspiration based on machine learning models, Big Earth Data, 9, 127–154, https://doi.org/10.1080/20964471.2024.2423431, 2025.

Yang, Y., Long, D., and Shang, S.: Remote estimation of terrestrial evapotranspiration without using meteorological data, Geophys. Res. Lett., 40, 3026–3030, https://doi.org/10.1002/grl.50450, 2013.

Yu, L., Qiu, G. Y., Yan, C., Zhao, W., Zou, Z., Ding, J., Qin, L., and Xiong, Y.: A global terrestrial evapotranspiration product based on the three-temperature model with fewer input parameters and no calibration requirement, Earth Syst. Sci. Data, 14, 3673–3691, https://doi.org/10.5194/essd-14-3673-2022, 2022.

Zeng, Z., Peng, L., and Piao, S.: Response of terrestrial evapotranspiration to Earth's greening, Curr. Opin. Environ. Sustain., 33, 9–25, https://doi.org/10.1016/j.cosust.2018.03.001, 2018.

Zhang, K., Zhu, G., Ma, N., Chen, H., and Shang, S.: Improvement of evapotranspiration simulation in a physically based ecohydrological model for the groundwater–soil–plant–atmosphere continuum, J. Hydrol., 613, 128440, https://doi.org/10.1016/j.jhydrol.2022.128440, 2022.

Zhang, L. L., Vrieling, A., Marshall, M., and Nelson, A.: A macroscale evapotranspiration benchmark based on spatial and temporal enhancements to the Budyko framework, J. Hydrol., 661, 133625, https://doi.org/10.1016/j.jhydrol.2025.133625, 2025.

Zhang, X., Zhang, X., Terfa, B. K., Nam, W, H., Zeng, J. Y., Ma, H. L., Gu, X. H., Du, W. Y., Wang, C., Yang, J., Wang, P., Niyogi, D., and Chen, N. C.: Mapping global drought-induced forest mortality based on multiple satellite vegetation optical depth data, Remote Sens. Environ., 315, 114406, https://doi.org/10.1016/j.rse.2024.114406, 2024.

Zhao, W. R., Jin, J. L., Bao, Z. X., Wu, J. R., Zhang, Q. X., Yu, C., and Wang, G. Q.: Direct and indirect effects of climatic factors on ecosystem carbon and water fluxes and water use efficiency across different climatic zones in China, Ecol. Indic., 175, 113565, https://doi.org/10.1016/j.ecolind.2025.113565, 2025.

Zhu, S. Y., Clement, R., McCalmont, J., Davies, C. A., and Hill, T.: Stable gap-filling for longer eddy covariance data gaps: A globally validated machine-learning approach for carbon dioxide, water, and energy fluxes, Agric. For. Meteorol., 314, 108777, https://doi.org/10.1016/j.agrformet.2021.108777, 2022.

Zou, L., Zhan, C., Xia, J., Wang, T., and Gippel, C. J.: Implementation of evapotranspiration data assimilation with catchment scale distributed hydrological model via an ensemble Kalman filter, J. Hydrol., 549, 685–702, https://doi.org/10.1016/j.jhydrol.2017.04.036, 2017.